PATENT APPLICATION

GENES FROM THE 20q13 AMPLICON AND THEIR USES

Inventors:

Joe W. Gray
Colin Collins
Soo-in Hwang
Tony Godfrey
David Kowbel
Johanna Rommens

Assignee:

Regents of the University of California

Entity:

TOWNSEND and TOWNSEND and CREW
Two Embarcadero Center, 8th Floor
San Francisco, CA, 94111-3834

GENES FROM 20q13 AMPLICON AND THEIR USES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of USSN 08/680,395 filed on July 15, 1996 which is related to copending U.S. Patent Application, USSN 08/546,130, filed October 20, 1995, both of which are incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

This invention pertains to the field to the field of cytogenetics. More particularly this invention pertains to the identification of genes in a region of amplification at about 20q13 in various cancers. The genes disclosed here can be used as probes specific for the 20q13 amplicon as well as for treatment of various cancers.

Chromosome abnormalities are often associated with genetic disorders, degenerative diseases, and cancer. In particular, the deletion or multiplication of copies of whole chromosomes or chromosomal segments, and higher level amplifications of specific regions of the genome are common occurrences in cancer. See, for example Smith, *et al.*, *Breast Cancer Res. Treat.*, 18: Suppl. 1: 5-14 (1991, van de Vijer & Nusse, *Biochim. Biophys. Acta.* 1072: 33-50 (1991), Sato, *et al.*, *Cancer. Res.*, 50: 7184-7189 (1990). In fact, the amplification and deletion of DNA sequences containing proto-oncogenes and tumor-suppressor genes, respectively, are frequently characteristic of tumorigenesis. Dutrillaux, *et al.*, *Cancer Genet. Cytogenet.*, 49: 203-217 (1990). Clearly the identification of amplified and deleted regions and the cloning of the genes involved is crucial both to the study of tumorigenesis and to the development of cancer diagnostics.

The detection of amplified or deleted chromosomal regions has traditionally been done by cytogenetics. Because of the complex packing of DNA into the chromosomes, resolution of cytogenetic techniques has been limited to regions larger than about 10 Mb; approximately the width of a band in Giemsa-stained chromosomes. In complex karyotypes with multiple translocations and other genetic changes, traditional cytogenetic analysis is of little utility because karyotype information is lacking or cannot be interpreted. Teyssier, J.R., *Cancer Genet. Cytogenet.*, 37: 103 (1989). Furthermore

conventional cytogenetic banding analysis is time consuming, labor intensive, and frequently difficult or impossible.

More recently, cloned probes have been used to assess the amount of a given DNA sequence in a chromosome by Southern blotting. This method is effective even if the genome is heavily rearranged so as to eliminate useful karyotype information. However, Southern blotting only gives a rough estimate of the copy number of a DNA sequence, and does not give any information about the localization of that sequence within the chromosome.

Comparative genomic hybridization (CGH) is a more recent approach to identify the presence and localization of amplified/deleted sequences. See Kallioniemi, *et al., Science,* 258: 818 (1992). CGH, like Southern blotting, reveals amplifications and deletions irrespective of genome rearrangement. Additionally, CGH provides a more quantitative estimate of copy number than Souther blotting, and moreover also provides information of the localization of the amplified or deleted sequence in the normal chromosome.

Using CGH, the chromosomal 20q13 region has been identified as a region that is frequently amplified in cancers (*see, e.g.* U.S. Patent No.). Initial analysis of this region in breast cancer cell lines identified a region approximately 2 Mb on chromosome 20 that is consistently amplified.

SUMMARY OF THE INVENTION

The present invention relates to the identification of a narrow region (about 600 kb) within a 2 Mb amplicon located at about chromosome 20q13 (more precisely at 20q13.2) that is consistently amplified in primary tumors. In addition this invention provides cDNA sequences from a number of genes which map to this region. Also provided is a contig (a series of clones that contiguously spans this amplicon) which can be used to prepare probes specific for the amplicon. The probes can be used to detect chromosomal abnormalities at 20q13.

Thus, in one embodiment, this invention provides a method of detecting a chromosome abnormality (*e.g.*, an amplification or a deletion) at about position FLpter 0.825 on human chromosome 20 (20q13.2). The method involves contacting a chromosome sample from a patient with a composition consisting essentially of one or

more labeled nucleic acid probes each of which binds selectively to a target polynucleotide sequence at about position FLpter 0.825 on human chromosome 20 under conditions in which the probe forms a stable hybridization complex with the target sequence; and detecting the hybridization complex. The step of detecting the hybridization complex can

5   involve determining the copy number of the target sequence. The probe preferably comprises a nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the nucleic acids disclosed here. Even more preferably, the probe comprises a subsequence selected from sequences set forth in SEQ. ID. Nos. 1-10 and 12. The probe is preferably labeled, and is more preferably labeled with digoxigenin or biotin.

10   In one embodiment, the hybridization complex is detected in interphase nuclei in the sample. Detection is preferably carried out by detecting a fluorescent label (*e.g.*, FITC, fluorescein, or Texas Red). The method can further involve contacting the sample with a reference probe which binds selectively to a chromosome 20 centromere.

This invention also provides for two new genes ZABC1 and 1b1 in the

15   20q13.2 region that are both amplified and overexpressed in a variety of cancers. ZABC1 appears to be a zinc finger protein containing a number of transcription factors that are expected interfere with normal transcription in cells in which they are overexpressed. ZABC1 and 1b1 thus appear to play an important role in the etiology of a number of cancers.

20   This invention also provides for proteins encoded by nucleic acid sequences in the 20q13 amplicon (SEQ. ID. Nos: 1-10 and 12) and subsequences more preferably subsequences of at least 10 amino acids, preferably of at least 20 amino acids and most preferably of at least 30 amino acids in length. Particularly preferred subsequences are epitopes specific to the 20q13 proteins more preferably epitopes specific to the ZABC1 and

25   1b1 proteins. Such proteins include, but are not limited to isolated polypeptides comprising at least 20 amino acids from a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 wherein the polypeptide, when presented as an immunogen, elicits the production of an antibody which specifically binds to a polypeptide selected from the group consisting of a polypeptide

30   encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 and the polypeptide does not bind to antiseral raised against a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ.

ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 which has been fully immunosorbed with a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11.

5   In another embodiment, the method can involve detecting a polypeptide (protein) encoded by a nucleic acid (ORF) in the 20q13 amplicon. The method may include any of a number of well known protein detection methods including, but not limited to, the protein assays disclosed herein.

     This invention also provides cDNA sequences from genes in the amplicon

10 (SEQ. ID. Nos. 1-10 and 12). [The nucleic acid sequences can be used in therapeutic applications according to known methods for modulating the expression of the endogenous gene or the activity of the gene product. Examples of therapeutic approaches include, antisense inhibition of gene expression, gene therapy, monoclonal antibodies that specifically bind the gene products, and the like. The genes can also be used for

15 recombinant expression of the gene products *in vitro*.

     This invention also provides for proteins (*e.g.,* SEQ. ID. No. 11) encoded by the cDNA sequences from genes in the amplicon (*e.g.,* SEQ. ID. Nos. 1-10 and 12). Where the amplified nucleic acids include cDNA which are expressed, detection and/or quantification of the protein expression product can be used to identify the presence or

20 absence or quantify the amplification level of the amplicon or of abnormal protein products produced by the amplicon.

     The probes disclosed here can be used in kits for the detection of a chromosomal abnormality at about position FLpter 0.825 on human chromosome 20. The kits include a compartment which contains a labeled nucleic acid probe which binds

25 selectively to a target polynucleotide sequence at about FLpter 0.825 on human chromosome 20. The probe preferably includes at least one nucleic acid that specifically hybridizes under stringent conditions to a nucleic acid selected from the nucleic acids disclosed here. Even more preferably, the probes comprise one or more nucleic acids selected from the nucleic acids disclosed here. In a preferred embodiment, the probes are

30 labelled with digoxigenin or biotin. The kit may further include a reference probe specific to a sequence in the centromere of chromosome 20.

Definitions

A "chromosome sample" as used herein refers to a tissue or cell sample prepared for standard *in situ* hybridization methods described below. The sample is prepared such that individual chromosomes remain substantially intact and typically comprises metaphase spreads or interphase nuclei prepared according to standard techniques.

"Nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, would encompass known analogs of natural nucleotides that can function in a similar manner as naturally occurring nucleotides.

An "isolated" polynucleotide is a polynucleotide which is substantially separated from other contaminants that naturally accompany it, e.g., protein, lipids, and other polynucleotide sequences. The term embraces polynucleotide sequences which have been removed or purified from their naturally-occurring environment or clone library, and include recombinant or cloned DNA isolates and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

"Subsequence" refers to a sequence of nucleic acids that comprise a part of a longer sequence of nucleic acids.

A "probe" or a "nucleic acid probe", as used herein, is defined to be a collection of one or more nucleic acid fragments whose hybridization to a target can be detected. The probe is labeled as described below so that its binding to the target can be detected. The probe is produced from a source of nucleic acids from one or more particular (preselected) portions of the genome, for example one or more clones, an isolated whole chromosome or chromosome fragment, or a collection of polymerase chain reaction (PCR) amplification products. The probes of the present invention are produced from nucleic acids found in the 20q13 amplicon as described herein. The probe may be processed in some manner, for example, by blocking or removal of repetitive nucleic acids or enrichment with unique nucleic acids. Thus the word "probe" may be used herein to refer not only to the detectable nucleic acids, but to the detectable nucleic acids in the form in which they are applied to the target, for example, with the blocking nucleic acids, *etc.* The blocking nucleic acid may also be referred to separately. What "probe" refers to specifically is clear from the context in which the word is used.

"Hybridizing" refers the binding of two single stranded nucleic acids via complementary base pairing.

"Bind(s) substantially" or "binds specifically" or "binds selectively" or "hybridizes specifically" refer to complementary hybridization between an oligonucleotide and a target sequence and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence. These terms also refer to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Typically, stringent conditions will be those in which the salt concentration is at least about 0.02 Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 60°C for short probes. Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

One of skill will recognize that the precise sequence of the particular probes described herein can be modified to a certain degree to produce probes that are "substantially identical" to the disclosed probes, but retain the ability to bind substantially to the target sequences. Such modifications are specifically covered by reference to the individual probes herein. The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 90% sequence identity, more preferably at least 95%, compared to a reference sequence using the methods described below using standard parameters.

Two nucleic acid sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the

complementary sequence is identical to all or a portion of a reference polynucleotide sequence.

Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two sequences over a "comparison window" to

5   identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

10   Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), by computerized implementations of these algorithms.

15   "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the

20   number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

Another indication that nucleotide sequences are substantially identical is if

25   two molecules hybridize to the same sequence under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence

30   hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar or lower at pH 7 and the temperature is at least about 60°C.

The term "20q13 amplicon protein" is used herein to refer to proteins encoded ORFs in the 20q13 amplicon disclosed herein. Assays that detect 20q13 amplicon proteins are are intended to detect the level of endogenous (native) 20q13 amplicon proteins present in subject biological sample. However, exogenous 20q13 amplicon

5 proteins (from a source extrinsic to the biological sample) may be added to various assays to provide a label or to compete with the native 20q13 amplicon protein in binding to an anti-20q13 amplicon protein antibody. One of skill will appreciate that a 20q13 amplicon protein mimetic may be used in place of exogenous 20q13 protein in this context. A "20q13 protein", as used herein, refers to a molecule that bears one or more 20q13

10 amplicon protein epitopes such that it is specifically bound by an antibody that specifically binds a native 20q13 amplicon protein.

As used herein, an "antibody" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda,

15 alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

The basic immunoglobulin (antibody) structural unit is known to comprise a

20 tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

25 Antibodies may exist as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'$_2$ a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab)'$_2$ may be reduced under mild conditions to break the disulfide linkage in

30 the hinge region thereby converting the F(ab)'$_2$ dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, *Fundamental Immunology*, W.E. Paul, ed., Raven Press, N.Y. (1993) for a more detailed description of

other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein also includes antibody fragments either produced

5   by the modification of whole antibodies or synthesized *de novo* using recombinant DNA methodologies.

The phrase "specifically binds to a protein" or "specifically immunoreactive with", when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and

10   other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies can be raised to the a 20q13 amplicon protein that bind the 20q13 amplicon protein and not

15   to any other proteins present in a biological sample. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual,* Cold Spring Harbor Publications, New York, for a

20   description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1(A) shows disease-free survival of 129 breast cancer patients

25   according to the level of 20q13 amplification. Patients with tumors having high level 20q13 amplification have a shorter disease-free survival (p=0.04 by Mantel-Cox test) compared to those having no or low level amplification.

Figure 1(B) Shows the same disease-free survival difference of Figure 4(A) in the sub-group of 79 axillary node-negative patients (p=0.0022 by Mantel-Cox test).

30   Figure 2 shows a comparison of 20q13 amplification detected by FISH in a primary breast carcinoma and its metastasis from a 29-year patient. A low level amplification of 20q13 (20q13 compared to 20p reference probe) was found in the primary

tumor. The metastasis, which appeared 8 months after mastectomy, shows a high level amplification of the chromosome 20q13 region. The overall copy number of chromosome 20 (20p reference probe) remained unchanged. Each data point represents gene copy numbers in individual tumor cells analyzed.

5   Figure 3 shows a graphical representation of the molecular cytogenetic mapping and subsequent cloning of the 20q13.2 amplicon. Genetic distance is indicated in centiMorgans (cM). The thick black bar represents the region of highest level amplification in the breast cancer cell line BT474 and covers a region of about 1.5 Mb. P1 and BAC clones are represented as short horizontal lines and YAC clones as heavier

10   horizontal lines. Not all YAC and P1 clones are shown. YACs 957f3, 782c9, 931h12, and 902 are truncated. Sequence tagged sites (STSs) appear as thin vertical lines and open circles indicate that a given YAC has been tested for and is positive for a given STS. Not all STSs have been tested on all YACs. The interval from which more than 100 exons have been trapped is represented as a filled box. The 600 kb interval spanning the region

15   of highest amplification level in primary tumors is represented by the filled black box (labeled Sequence). The lower part of the figure shows the levels of amplification in two primary tumors that further narrow the region of highest amplification to within about 600 kb.

          Figure 4 provides a higher resolution map of the amplicon core as defined in

20   primary tumors.

          Figure 5 shows the map location of 15 genes in the amplicon.

DETAILED DESCRIPTION

25   This invention provides a number of cDNA sequences which can be used as probes for the detection of chromosomal abnormalities at 20q13. Studies using comparative genomic hybridization (CGH) have shown that a region at chromosome 20q13 is increased in copy number frequently in cancers of the breast (~30%), ovary (~ 15%), bladder (~30%), head and neck (~75%) and colon (~80%). This suggests the presence of

30   one or more genes that contribute to the progression of several solid tumors are located at 20q13.

Gene amplification is one mechanism by which dominantly acting oncogenes are overexpressed, allowing tumors to acquire novel growth characteristics and/or resistance to chemotherapeutic agents. Loci implicated in human breast cancer progression and amplified in 10-25% of primary breast carcinomas include the erbB-2 locus (Lupu *et al., Breast Cancer Res. Treat.*, 27: 83 (1993), Slamon *et al. Science*, 235: 177-182 (1987), Heiskanen *et al. Biotechniques*, 17: 928 (1994)) at 17q12, cyclin-D (Mahadevan *et al., Science*, 255: 1253-1255 (1993), Gillett *et al., Canc. Res.*, 54: 1812 (1994)) at 11q13 and MYC (Gaffey *et al., Mod. Pathol.*, 6: 654 (1993)) at 8q34.

Pangenomic surveys using comparative genomic hybridization (CGH) recently identified about 20 novel regions of increased copy number in breast cancer (Kallioniemi *et al., Genomics,* 20: 125-128 (1994)). One of these loci, band 20q13, was amplified in 18% of primary tumors and 40% of cell lines (Kallioniemi *et al., Genomics,* 20: 125-128 (1994)). More recently, this same region was found amplified in 15% of ovarian, 80% of bladder and 80% of colorectal tumors. The resolution of CGH is limited to 5-10 Mb. Thus, FISH was performed using locus specific probes to confirm the CGH data and precisely map the region of amplification.

The 20q13 region has been analyzed in breast cancer at the molecular level and a region, approximately 600 kb wide, that is consistently amplified was identified, as described herein. Moreover, as shown herein, the importance of this amplification in breast cancer is indicated by the strong association between amplification and decreased patient survival and increased tumor proliferation (specifically, increased fraction of cells in S-phase).

In particular, as explained in detail in Example 1, high-level 20q13 amplification was associated (p=0.0022) with poor disease free survival in node-negative patients, compared to cases with no or low-level amplification (Figure 1). Survival of patients with moderately amplified tumors did not differ significantly from those without amplification. Without being bound to a particular theory, it is suggested that an explanation for this observation may be that low level amplification precedes high level amplification. In this regard, it may be significant that one patient developed a local metastasis with high-level 20q13.2 amplification 8 month after resection of a primary tumor with low level amplification (Figure 3).

The 20q13 amplification was associated with high histologic grade of the tumors. This correlation was seen in both moderately and highly amplified tumors. There was also a correlation (p=0.0085) between high level amplification of a region complementary to a particular probe, RMC20C001 (Tanner *et al., Cancer Res.*, 54: 4257-4260 (1994)), and cell proliferation, measured by the fraction of cells in S-phase (Figure 4). This finding is important because it identifies a phenotype that can be scored in functional assays, without knowing the mechanism underlying the increased S-phase fraction. The 20q13 amplification did not correlate with the age of the patient, primary tumor size, axillary nodal or steroid hormone-receptor status.

This work localized the 20q13.2 amplicon to an interval of approximately 2 Mb. Furthermore, it suggests that high-level amplification, found in 7% of the tumors, confers an aggressive phenotype on the tumor, adversely affecting clinical outcome. Low level amplification (22% of primary tumors) was associated with pathological features typical of aggressive tumors (high histologic grade, aneuploidy and cell proliferation) but not patient prognosis.

In addition, it is shown herein that the 20q13 amplicon (more precisely the 20q13.2 amplicon) is one of three separate co-amplified loci on human chromosome 20 that are packaged together throughout the genomes of some primary tumors and breast cancer cell lines. No known oncogenes map in the 20q13.2 amplicon.

Identification of 20q13 Amplicon Probes.

Initially, a paucity of available molecular cytogenetic probes dictated that FISH probes be generated by the random selection of cosmids from a chromosome 20 specific library, LA20NC01, and map them to chromosome 20 by digital imaging microscopy. Approximately 46 cosmids, spanning the 70 Mb chromosome, were isolated for which fractional length measurements (FLpter) and band assignments were obtained. Twenty six of the cosmids were used to assay copy number in the breast cancer cell line BT474 by interphase FISH analysis. Copy number was determined by counting hybridization signals in interphase nuclei. This analysis revealed that cosmid RMC20C001 (Flpter, 0.824; 20q13.2), described by Stokke *et al., Genomics*, 26: 134-137 (1995), defined the highest-level amplification (~60 copies/cell) in BT474 cells (Tanner *et al., Cancer Res.*, 54: 4257-4260 (1994)).

P1 clones containing genetically mapped sequences were selected from 20q13.2 and used as FISH probes to confirm and further define the region of amplification. Other P1 clones were selected for candidate oncogenes broadly localized to the 20q13.2 region (Flpter, 0.81-0.84). These were selected from the DuPont P1 library

5 (Shepherd, *et al.*, *Proc. Natl. Acad. Sci. USA*, 92: 2629 (1994), available commercially from Genome Systems), by PCR (Saiki *et al*, *Science*, 230: 1350 (1985)) using primer pairs developed in the 3' untranslated region of each candidate gene. Gene specific P1 clones were obtained for, protein tyrosine phosphatase (PTPN1, Flpter 0.78), melanocortin 3 receptor (MC3R, Flpter 0.81), phosphoenolpyruvate carboxy kinase (PCK1, Flpter

10 0.85), zinc finger protein 8 (ZNF8, Flpter 0.93), guanine nucleotide-binding protein (GNAS 1, Flpter .873), src-oncogene (SRC, Flpter 0.669), topoisomerase 1 (TOP1, Flpter 0.675), the bcl-2 related gene bcl-x (Flpter 0.526) and the transcription factor E2F-1 (FLpter 0.541). Each clone was mapped by digital imaging microscopy and assigned Flpter values. Five of these genes (SRC, TOPO1, GNAS1, E2F-1 and BC1-x) were

15 excluded as candidate oncogenes in the amplicon because they mapped well outside the critical region at Flpter 0.81-0.84. Three genes (PTPNR1, PCK-1 and MC3R) localized close enough to the critical region to warrant further investigation.

Interphase FISH on 14 breast cancer cell lines and 36 primary tumors using 24 cosmid and 3 gene specific P1 (PTPNRL, PCK-1 and MC3R) probes found high level

20 amplification in 35% (5/14) of breast cancer cell lines and 8% (3/36) of primary tumors with one or more probe. The region with the highest copy number in 4/5 of the cell lines and 3/3 primary tumors was defined by the cosmid RMC20C001. This indicated that PTPNR1, PCK1 and MC3R could also be excluded as candidates for oncogenes in the amplicon and, moreover, narrowed the critical region from 10 Mb to 1.5-2.0 Mb (*see,*

25 Tanner *et al.*, *Cancer Res.*, 54: 4257-4260 (1994).

Because probe RMC20C001 detected high-level (3 to 10-fold) 20q13.2 amplification in 35% of cell lines and 8% of primary tumors it was used to (1) define the prevalence of amplification in an expanded tumor population, (2) assess the frequency and level of amplification in these tumors, (3) evaluate the association of the 20q13.2 amplicon

30 with pathological and biological features, (4) determine if a relationship exists between 20q13 amplification and clinical outcome and (5) assess 20q13 amplification in metastatic breast tumors.

As detailed in Example 1, fluorescent *in situ* hybridization (FISH) with RMC20C001 was used to assess 20q13.2 amplification in 132 primary and 11 recurrent breast tumors. The absolute copy number (mean number of hybridization signals per cell) and the level of amplification (mean number of signals relative to the p-arm reference probe) were determined. Two types of amplification were found: Low level amplification (1.5-3 fold with FISH signals dispersed throughout the tumor nuclei) and high level amplification (>3 fold with tightly clustered FISH signals). Low level 20q13.2 amplification was found in 29 of the 132 primary tumors (22%), whereas nine cases (6.8%) showed high level amplification.

RMC20C001 and four flanking P1 probes (MC3R, PCK, RMC20C026, and RMC20C030) were used to study the extent of DNA amplification in highly amplified tumors. Only RMC20C001 was consistently amplified in all tumors. This finding confirmed that the region of common amplification is within a 2 Mb interval flanked by but not including PCK-1 and MC3R.

A physical map was assembled to further localize the minimum common region of amplification and to isolate the postulated oncogene(s). The DuPont P1 library (Shepherd *et al. Proc. Natl. Acad. Sci. USA*, 91: 2629 (1994) was screened for STSs likely to map in band 20q13.2. P1 clones at the loci D20S102, D20S100, D20S120, D20S183, D20S480, D20S211 were isolated, and FISH localized each to 20q13.2. Interphase FISH analysis was then performed in the breast cancer cell line BT474 to assess the amplification level at each locus. The loci D20S100-D20S120-D20S183-D20S480-D20S211 were highly amplified in the BT474 cell line, whereas D20S102 detected only low level amplification. Therefore, 5 STSs, spanning 5 cM, were localized within the 20q13.2 amplicon and were utilized to screen the CEPH megaYAC library.

CEPH megaYAC library screening and computer searches of public databases revealed D20S120-D20S183-D20S480-D20S211 to be linked on each of three megaYAC clones y820f5, 773h10, and 931h6 (Figure 3). Moreover, screening the CEPH megaYAC library with STSs generated from the ends of cosmids RMC20C001, RMC20C30 and RMC20C028 localized RMC20C001 to each of the same three YAC clones. It was estimated, based on the size of the smallest of these YAC clones, that D20S120-D20S 183-RMC20C001-D20S480-D20S211 map into an interval of less than 1.1 Mb. D20S100 was localized 300 kb distal to D20S120 by interphase FISH and to

YAC901b12 by STS mapping. The combined STS data made it possible to construct a 12 member YAC contig which spans roughly 4 Mb encompassing the 1.5 Mb amplicon and containing the loci RMC20C030-PCK1-RMC20C001-MC3R-RMC20CO26. Each YAC was mapped by FISH to confirm localization to 20q13.2 and to check for chimerism. Five

5   clonal isolates of each YAC were sized by pulsed field gel electrophoresis (PFGE). None of the YACs are chimeric, however, several are highly unstable.

The YAC contig served as a framework from which to construct a 2 MbP1 contig spanning the 20Q13 amplicon. P1 clones offered numerous advantages over YAC clones including (1) stability, (2) a chimeric frequency of less than 1%, (3) DNA isolation

10   by standard miniprep procedures, (4) they make ideal FISH probes, (5) the ends can be sequenced directly, (6) engineered γδ transposons carrying bidirectional primer binding sites can be integrated at any position in the cloned DNA (Strathmann *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 1247 (1991)) (7) P1 clones are the templates for sequencing the human and Drosophila genomes at the LBNL HGC (Palazzolo *et al. DOE Human Genome*

15   *Program, Contractor-Grantee Workshop IV. Santa Fe, New Mexico*, November 13-17 1994).

About 90 P1 clones were isolated by screening the DuPont P1 library either by PCR or filter hybridization. For PCR based screening, more than 22 novel STSs were created by two methods. In the first method, the ends of P1 clones localized to the

20   amplicon were sequenced, STSs developed, and the P1 library screened for walking clones. In the second approach inter-Alu PCR (Nelson *et al.*, 86: 6686-6690 (1989)) was performed on YACs spanning the amplicon and the products cloned and sequenced for STS creation. In the filter based hybridization scheme P1 clones were obtained by performing inter-Alu PCR on YACs spanning the amplicon, radio-labeling the products and

25   hybridizing these against filters containing a gridded array of the P1 library. Finally, to close gaps a human genomic bacterial artificial chromosome (BAC) library (Shizuya *et al. Proc. Natl. Acad. Sci. USA*, 89: 8794 (1992), commercially available from Research Genetics, Huntsville, Alabama, USA) was screened by PCR. These methods combined to produce more than 100 P1 and BAC clones were localized to 20q13.2 by FISH. STS

30   content mapping, fingerprinting, and free-chromatin fish (Heiskanen *et al., BioTechniques*, 17: 928 (1994)) were used to construct the 2 Mb contig shown in Figure 3.

Fine Mapping the 20q13.2 Amplicon in BT474

Clones from the 2 Mb P1 contig were used with FISH to map the level of amplification at 20q13.2 in the breast cancer cell line BT474. 35 P1 probes distributed at regular intervals along the contig were used. The resulting data indicated that the region of *highest copy number increase* in BT474 occurs between D20S100 and D20S211, an interval of approximately 1.5 Mb. P1 FISH probes, in this interval, detect an average of 50 signals per interphase nuclei in BT474, while no, or only low level amplification, was detected with the P1 clones outside this region. Thus, both the proximal and distal boundaries of the amplicon were cloned.

Fine Mapping the 20q13.2 Amplicon in Primary Tumors.

Fine mapping the amplicon in primary tumors revealed the *minimum common region of high amplification* that is of pathobiological significance. This process is analogous to screening for informative meiosis in the narrowing of genetic intervals encoding heritable disease genes. Analysis of 132 primary tumors revealed thirty-eight primary tumors that are amplified at the RMC20C001 locus. Nine of these tumors have high level amplification at the RMC20C001 locus and were further analyzed by interphase FISH with 8 P1s that span the ≈2 Mb contig. The minimum common region of amplification was mapped to a ≈600 kb interval flanked by P1 clones #3 and #12 with the highest level of amplification detected by P1 clone #38 corresponding to RMC20C001 (Figure 4).

The P1 and BAC clones spanning the 600 kb interval of the 20q13 amplicon are listed in Table 1 which provides a cross-reference to the DuPont P1 library described by Shepherd, *et al., Proc. Natl. Acad. Sci. USA*, 92: 2629 (1994). These P1 and BAC probes are available commercially from Genetic Systems, and Research Genetics, respectively).

Table 1. Cross reference to probes of the DuPont P1 library (Shepherd, *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:2629 (1994) which is commerically available from Genomic Systems, St. Louis, Missouri, USA). PCR primers are illustrated for amplication of Sequence tag sites for each clone. In addition, PCR conditions (Mg concentration and annealing temperature), as well as PCR product size, is provided. Size: PCR product Size; mM MgCl: Mg concentration; Ann.: Annealing temperature; P1: P1 probe ID number; PC: DuPont Library Plate Coordinates; SCA: DuPont Library Single Clone address; SEQ-forward and SEQ-backward: forward and backward PCR primers, respectively; SEQ ID NO:-forward and SEQ ID NO:-backward: Sequence Listing SEQ ID NO: for forward and backward primers, respectively.

PRIMER NAME	SIZE (bp)	mM MgCl	Ann.	P1	PC	SCA	SEQ-forward	SEQ-backward	SEQ ID NO:-forward	SEQ ID NO:-backward
352.32	136	1.5	52	20	103e5	1228e	TTGGCATTGGTATCAGGTAGCTG	TTGGAGCAGAGAGGGGATTGTGTG	9/13	10/14
388.13 F1/B1	201	1.5	52	17	69g6	821	AATCCCCTCAAACCCTGCTGCTAC	TGGAGCCTGAACTTCTGCAATC	11/15	12/16
				15	98f4	1167f				
D20S183	270	3	48	30	24h1	124g6	CCGGGATACCGACATTG	TGCACATAAAACAGCCAGC	13/17	14/18
				40		276h				
D20S211 1/2	135	1.5	52	29	119f4	1418f	TTGGAATCAATGGAGCAAAA	AGTCTTACCCAATGTGGTCC	15/19	16/20
D20S480	300	3	55	68	100d12	1199d2	GTGGTGAACACCACCAATAAATGG	AAGCAAATAAAACCAATAAACTCG	17/21	18/22
				41	86c1	1020c				
				42	103d9	1232d				
				67	91b2	1081b9				
9X-SP6 hmF/hmB	165	1.5	55	7	31d11	370d	CAAGATCTGACCCGTCAATC	GACTTCTTCAGGAAAGAGATCAGTG	19/23	20/24
				9	3519	416f				
11S-17 F2/B4	146	3	58	11	41b1	480b	GCCATGTACCCACCTGAAAAATC	TCAGAACACCCGTGCAGAATTAAG	21/25	22/26
12T-17 F2/B1	153	3	58	12	42c2	493c	CCTAAAACTTGGTGCTTAAATCTA	GTCTCACAAGGCAGATGTGG	23/27	24/28
28T F1/B1	219	1.5	52	74		888f2	TTTGTGTATGTTGAGCCATC	CTTCCAATCTCATTCTATGAGG	25/29	26/30
				2	12c6	137c				
				25	118c11	1413c				
				26	118c11	1413c				
28S F1/B3	214	1.5	55	28	118g11	1413g	GCTTGTTTAAGTGTCACTAGGG	CACTCTGGTAAATGACCTTTGTC	27/31	28/32
				25	118c11	1413c				
				27	118g11	1413g				
				10	36f10	429f				
69S	100	3	55	69		412b5	CCTACACCATTCCAACTTTGG	GCCAGATGTATGTTTGCTACGGAAC	29/33	30/34
				5	23c1	264c				
HSCOFH032 F/B	129	1.5	55	3	12-e11	142e	TCTCAAACCTGTCCACTTCTTG	CTGCTGTGGTGGAGAATGG	31/35	32/36
				18	77a10	921a				
60A1	191	1.5	58	36	112g8	1139g	TGTCCTCCTTCTCCCTCATCCTAC	AATGCCTCCACTCACAGGAATG	33/37	34/38
				39	34a6	401a				
820F5A1T F1/B1	175	1.5	48	15	53c7	630c	CCTCTTCAGTGTCTTCCTATTGA	GGGAGGAGGTTGTAGGCAAC	35/39	36/40
				16	58b9	692b				

Table 2. Cross reference to probes of the BAC library. Clone # refers to the clone number provided, e.g., in Figure 4, while the plate coordinates are the plate coordinates in the Research Genetics BAC library. Size: PCR product Size; mM MgCl: Mg concentration; Ann.: Annealing temperature; BAC #: BAC probe ID number; SEQ-forward and SEQ-backward: forward and backward PCR primers, respectively; SEQ ID NO:-forward and SEQ ID NO:-backward: Sequence Listing SEQ ID NO: for forward and backward primers, respectively.

PRIMER NAME	SIZE (bp)	mM MgCl	Ann.	BAC #	BAC Plate Coordinates	SEQ-forward	SEQ-backward	SEQ ID NO:-forward	SEQ ID NO:-backward
18T F1/B1	156	3	62	99	L11 plate 146	AGCAAAGCAAAGGTGGCACAC	TGACATGGGAGAGAGACACACTTCC	37 41	38 42
9S F1/B1	214	1.5	55	97	E8 plate 183	AGGTTTACCAATGTGTTGG	TCTACATCCCATTCTCTCTG	39 43	40 44
D20S480	300	3	55	95	H15 plate 140	GTGGTGAACACCAATAAATGG	AAGCAAATAAAACCAATAAACTCG	17 21	18 22
D20S211 1/2	135	1.5	52	103	A15 plate 188	TTGGAATCAATGGAGCAAAA	AGCTTTACCCAATGTGGTCC	18 19	16 20
				102	A1 plate 46	.	.	.	
11S-17 F2/B4	146	3	58	100	E4 plate 43	GCCATGTACCCACCTGAAAAATC	TCAGAACACCCGTGCAGAATTAAG	24 25	22 26
				101	J5 plate 118	.	.		

cDNA sequences from the 20q13 amplicon.

Exon trapping (see, e.g., Duyk et al., Proc. Natl. Acad. Sci. USA, 87: 8995-8999 (1990) and Church et al., Nature Genetics, 6: 98-105 (1994)) was performed on the P1 and BAC clones spanning the ≈600 kb minimum common region of amplification and has isolated more than 200 exons.

Analysis of the exons DNA sequence revealed a number of sequence similarities (85% to 96%) to partial CDNA sequences in the expressed sequence data base (dbest) and to a S. cerevisiae chromosome XIV open reading frame. Each P1 clone subjected to exon trapping has produced multiple exons consistent with at least a medium density of genes. Over 200 exons have been trapped and analyzed as well as 200 clones isolated by direct selection from a BT474 cDNA library. In addition a 0.6 Mb genomic interval spanning the minimal amplicon described below is being sequenced. Exon prediction and gene modeling are carried out with XGRAIL, SORFIND, and BLAST programs. Gene fragments identified by these approaches have been analyzed by RT-PCR, Northern and Southern blots. Fifteen unique genes were identifed in this way (see, Table 3 and Figure 5).

In addition two other genes ZABC1 (SEQ. ID. 9 and 10) and 1b1 (SEQ ID No. 12) were also were shown to be overexpressed in a variety of different cancer cells.

Sequence information from various cDNA clones are provided below. They are as follows:

3bf4 (SEQ. ID. No. 1) - 3kb transcript with sequence identity to a tyrosine kinase gene, termed A6, disclosed in Beeler et al. Mol. Cell. Biol. 14:982-988 (1994) and WO 95/19439. These references, however, do not disclose that the gene is located in the chromosome 20 amplicon.

1b11 (SEQ. ID. No. 2) - an approximately 3.5 kb transcript whose expression shows high correlation with the copy number of the amplicon. The sequence shows no homology with sequences in the databases searched.

cc49 (SEQ. ID. No. 3) - a 6-7 kb transcript which shows homology to C2H2 zinc finger genes.

cc43 (SEQ. ID. No. 4) - an approximately 4 kb transcript which is expressed in normal tissues, but whose expression in the breast cancer cell line has not been detected.

41.1 (SEQ. ID. No. 5) - shows homology to the homeobox T shirt gene in Drosophila.

GCAP (SEQ. ID. No. 6) - encodes a guanino cyclase activating protein which is involved in the biosynthesis of cyclic AMP. As explained in detail below, sequences from this gene can also be used for treatment of retinal degeneration.

1b4 (SEQ. ID. No. 7) - a serine threonine kinase.

20sa7 (SEQ. ID. No. 8) - a homolog of the rat gene, BEM-1.

In addition, the entire nucleotide sequence is provided for ZABC-1. ZABC-1 stands for zinc finger amplified in breast cancer. This gene maps to the core of the 20q13.2 amplicon and is overexpressed in primary tumors and breast cancer cell lines having 20q13.2 amplification. The genomic sequence (SEQ. ID. No. 9) includes roughly 2kb of the promoter region. SEQ ID. No. 10 provides the cDNA sequence derived open reading frame and SEQ ID. No. 11 provides the predicted protein sequence. Zinc finger containing genes are often transcription factors that function to modulate the expression of down stream genes. Several known oncogenes are infact zinc finger containing genes.

Finally, this invention also provides the full length cDNA sequence for a cDNA designated 1b1 (SEQ. ID. No. 12) which is is also overexpressed in numerous breast cancer cell lines and some primary tumors.

Table 3. Gene fragments identified by exon trapping and analyzed by RT-PCR, Northern and Southern blots.

Map #	Gene ID	Transcript Size (kb)	EST Identity (>95%)	Cloned Sequence (kb)	Protein Homologies	Map Location
1	20sa7	-	Yes	3	PTP	3, 18, 99
2	1b11	3.5	No	1.5	novel	18, 3, 99, 69
3	200.1	-	-	-	-	-
4	132.1	-	-	-	novel	132
5	132.2	-	-	-	-	132
6	3bf4	3	Yes	3	PTK	ambiguous
7	7.1	-	-	-	-	7, 11, 97
8	7.2	2.4	-	-	-	7, 11, 97
9	cc49	7, 4	Yes	3.6	Kruppel	97, 103
10	cc43	1.4	Yes	1.8	hypothetical yeast protein	97, 7, 11
11	et1807	2.5	Yes	0.7	novel	97, 9
12	et2106	None detected	Yes	-	-	95, 39, 38
13	41.1	7, 8, 11	Yes	3	hemeotic gene	95, 41, 42
14	67.1	-	Yes	2Kb	cGMP reg. protein	67
15	67.2	0	Yes	-	-	67

20q13 Amplicon Proteins

As indicated above, this invention also provides for proteins encoded by nucleic acid sequences in the 20q13 amplicon (*e.g.,* SEQ. ID. Nos: 1-10 and 12) and subsequences more preferably subsequences of at least 10 amino acids, preferably of at least 20 amino acids, and most preferably of at least 30 amino acids in length. Particularly preferred subsequences are epitopes specific to the 20q13 proteins more preferably epitopes specific to the ZABC1 and 1b1 proteins. Such proteins include, but are not limited to isolated polypeptides comprising at least 10 contiguous amino acids from a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 wherein the polypeptide, when presented as an immunogen, elicits the production of an antibody which specifically binds to a polypeptide selected from tehgroup consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 and the polypeptide does not bind to antiseral raised against a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11 which has been fully immunosorbed with a polypeptide selected from the group consisting of a polypeptide encoded by the nucleic acids of SEQ. ID No. 1-10 and 12 or from the polypeptide of SEQ. ID. No. 11.

A protein that specifically binds to or that is specifically immunoreactive with an antibody generated against a defined immunogen, such as an immunogen consisting of the amino acid sequence of SEQ ID NO 11 is determined in an immunoassay. The immunoassay uses a polyclonal antiserum which was raised to the protein of SEQ ID NO 11 (the immunogenic polypeptide). This antiserum is selected to have low crossreactivity against other similar known polypeptides and any such crossreactivity is removed by immunoabsorbtion prior to use in the immunoassay (*e.g.,* by immunosorbtion of the antisera with the related polypeptide).

In order to produce antisera for use in an immunoassay, the polypeptide *e.g.,* the polypeptide of SEQ ID NO 11 is isolated as described herein. For example, recombinant protein can be produced in a mammalian or other eukaryotic cell line. An inbred strain of mice is immunized with the protein of SEQ ID NO 11 using a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see Harlow and Lane, *supra*). Alternatively, a synthetic polypeptide derived from the

sequences disclosed herein and conjugated to a carrier protein is used as an immunogen. Polyclonal sera are collected and titered against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested

5   for their cross reactivity against known polypeptides using a competitive binding immunoassay such as the one described in Harlow and Lane, *supra*, at pages 570-573. Preferably more than one known polypeptide is used in this determination in conjunction with the immunogenic polypeptide.

The known polypeptides can be produced as recombinant proteins and

10   isolated using standard molecular biology and protein chemistry techniques as described herein.

Immunoassays in the competitive binding format are used for crossreactivity determinations. For example, the immunogenic polypeptide is immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the

15   immobilized antigen. The ability of the a proteins to compete with the binding of the antisera to the immobilized protein is compared to the immunogenic polypeptide. The percent crossreactivity for the protein is calculated, using standard calculations. Those antisera with less than 10% crossreactivity to known polypeptides are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by

20   immunoabsorbtion with known polypeptide.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described herein to compare a "target" polypeptide to the immunogenic polypeptide. To make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to

25   inhibit 50% of the binding of the antisera to the immobilized protein is determined using standard techniques. If the amount of the target polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the target polypeptide is said to specifically bind to an antibody generated to the immunogenic protein. As a final determination of specificity, the pooled antisera is fully immunosorbed with the

30   immunogenic polypeptide until no binding to the polypeptide used in the immunosorbtion is detectable. The fully immunosorbed antisera is then tested for reactivity with the test

polypeptide. If no reactivity is observed, then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

Similarly, in a reciprocal experiment, the pooled antisera is immusorbed with the test polypeptide. If the antisera which remains after the immusorbtion does not bind to the immunogenic polypeptide (*i.e.,* the polypeptide of SEQ ID NO: 11 used to elicit the antisera) then the test polypeptide is specifically bound by the antisera elicited by the immunogenic peptide.

Detection of 20q13 Abnormalities.

One of skill in the art will appreciate that the clones and sequence information provided herein can be used to detect amplifications, or other chromosomal abnormalities, at 20q13 in a biological sample. Generally the methods involve hybridization of probes that specifically bind one or more nucleic acid sequences of the target amplicon with nucleic acids present in a biological sample or derived from a biological sample.

As used herein, a biological sample is a sample of biological tissue or fluid containing cells desired to be screened for chromosomal abnormalities (*e.g.* amplifications of deletions). In a preferred embodiment, the biological sample is a cell or tissue suspected of being cancerous (transformed). Methods of isolating biological samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, needle biopsies, and the like. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. It will be recognized that the term "sample" also includes supernatant (containing cells) or the cells themselves from cell cultures, cells from tissue culture and other media in which it may be desirable to detect chromosomal abnormalities.

In a preferred embodiment, a biological sample is prepared by depositing cells, either as single cell suspensions or as tissue preparation, on solid supports such as glass slides and fixed by choosing a fixative which provides the best spatial resolution of the cells and the optimal hybridization efficiency.

Selecting Probes

Any of the P1 probes listed in Table 1, the BAC probes listed in Table 2, or the cDNAs disclosed here are suitable for use in detecting the 20q13 amplicon. Methods of preparing probes are well known to those of skill in the art (*see, e.g.* Sambrook *et.al., Molecular Cloning: A Laboratory Manual (2nd ed.)*, Vols. 1-3, Cold Spring Harbor Laboratory, (1989) or *Current Protocols in Molecular Biology,* F. Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987))

The probes are most easily prepared by combining and labeling one or more of the constructs listed in Tables 1 and 2. Prior to use, the constructs are fragmented to provide smaller nucleic acid fragments that easily penetrate the cell and hybridize to the target nucleic acid. Fragmentation can be by any of a number of methods well known to hose of skill in the art. Preferred methods include treatment with a restriction enzyme to selectively cleave the molecules, or alternatively to briefly heat the nucleic acids in the presence of Mg^{2+}. Probes are preferably fragmented to an average fragment length ranging from about 50 bp to about 2000 bp, more preferably from about 100 bp to about 1000 bp and most preferably from about 150 bp to about 500 bp.

Alternatively, probes can be produced by amplifying (*e.g.* via PCR) selected subsequences from the 20q13 amplicon disclosed herein. The sequences provided herein permit one of skill to select primers that amplify sequences from one or more exons located within the 20q13 amplicon.

Particularly preferred probes include nucleic acids from probes 38, 40, and 79, which corresponds to RMC20C001. In addition, the cDNAs are particularly useful for identifying cells that have increased expression of the corresponding genes, using for instance, Northern blot analysis.

One of skill will appreciate that using the sequence information and clones provided herein, one of skill in the art can isolate the same or similar probes from other human genomic libraries using routine methods (*e.g.* Southern or Northern Blots).

Labeling Probes

Methods of labeling nucleic acids are well known to those of skill in the art. Preferred labels are those that are suitable for use in *in situ* hybridization. The nucleic acid probes may be detectably labeled prior to the hybridization reaction. Alternatively, a

detectable label which binds to the hybridization product may be used. Such detectable labels include any material having a detectable physical or chemical property and have been well-developed in the field of immunoassays.

As used herein, a "label" is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Useful labels in the present invention include radioactive labels (e.g. ^{32}P, ^{125}I, ^{14}C, ^{3}H, and ^{35}S), fluorescent dyes (e.g. fluorescein, rhodamine, Texas Red, etc.), electron-dense reagents (e.g. gold), enzymes (as commonly used in an ELISA), colorimetric labels (e.g. colloidal gold), magnetic labels (e.g. Dynabeads™), and the like. Examples of labels which are not directly detected but are detected through the use of directly detectable label include biotin and dioxigenin as well as haptens and proteins for which labeled antisera or monoclonal antibodies are available.

The particular label used is not critical to the present invention, so long as it does not interfere with the in situ hybridization of the stain. However, stains directly labeled with fluorescent labels (e.g. fluorescein-12-dUTP, Texas Red-5-dUTP, etc.) are preferred for chromosome hybridization.

A direct labeled probe, as used herein, is a probe to which a detectable label is attached. Because the direct label is already attached to the probe, no subsequent steps are required to associate the probe with the detectable label. In contrast, an indirect labeled probe is one which bears a moiety to which a detectable label is subsequently bound, typically after the probe is hybridized with the target nucleic acid.

In addition the label must be detectible in as low copy number as possible thereby maximizing the sensitivity of the assay and yet be detectible above any background signal. Finally, a label must be chosen that provides a highly localized signal thereby providing a high degree of spatial resolution when physically mapping the stain against the chromosome. Particularly preferred fluorescent labels include fluorescein-12-dUTP and Texas Red-5-dUTP.

The labels may be coupled to the probes in a variety of means known to those of skill in the art. In a preferred embodiment the nucleic acid probes will be labeled using nick translation or random primer extension (Rigby, et al. J. Mol. Biol., 113: 237 (1977) or Sambrook, et al., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1985)).

One of skill in the art will appreciate that the probes of this invention need not be absolutely specific for the targeted 20q13 region of the genome. Rather, the probes are intended to produce "staining contrast". "Contrast" is quantified by the ratio of the probe intensity of the target region of the genome to that of the other portions of the genome. For example, a DNA library produced by cloning a particular chromosome (*e.g.* chromosome 7) can be used as a stain capable of staining the entire chromosome. The library contains both sequences found only on that chromosome, and sequences shared with other chromosomes. Roughly half the chromosomal DNA falls into each class. If hybridization of the whole library were capable of saturating all of the binding sites on the target chromosome, the target chromosome would be twice as bright (contrast ratio of 2) as the other chromosomes since it would contain signal from the both the specific and the shared sequences in the stain, whereas the other chromosomes would only be stained by the shared sequences. Thus, only a modest decrease in hybridization of the shared sequences in the stain would substantially enhance the contrast. Thus contaminating sequences which only hybridize to non-targeted sequences, for example, impurities in a library, can be tolerated in the stain to the extent that the sequences do not reduce the staining contrast below useful levels.

Detecting the 20q13 Amplicon.

As explained above, detection of amplification in the 20q13 amplicon is indicative of the presence and/or prognosis of a large number of cancers. These include, but are not limited to breast, ovary, bladder, head and neck, and colon.

In a preferred embodiment, a 20q13 amplification is detected through the hybridization of a probe of this invention to a target nucleic acid (*e.g.* a chromosomal sample) in which it is desired to screen for the amplification. Suitable hybridization formats are well known to those of skill in the art and include, but are not limited to, variations of Southern Blots, *in situ* hybridization and quantitative amplification methods such as quantitative PCR (*see, e.g.* Sambrook, *supra.,* Kallioniemi *et al., Proc. Natl Acad Sci USA,* 89: 5321-5325 (1992), and *PCR Protocols, A Guide to Methods and Applications,* Innis *et al.,* Academic Press, Inc. N.Y., (1990)).

In situ Hybridization.

In a preferred embodiment, the 20q13 amplicon is identified using *in situ* hybridization. Generally, *in situ* hybridization comprises the following major steps: (1) fixation of tissue or biological structure to analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target DNA, and to reduce nonspecific binding; (3) hybridization of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) posthybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagent used in each of these steps and their conditions for use vary depending on the particular application.

In some applications it is necessary to block the hybridization capacity of repetitive sequences. In this case, human genomic DNA is used as an agent to block such hybridization. The preferred size range is from about 200 bp to about 1000 bases, more preferably between about 400 to about 800 bp for double stranded, nick translated nucleic acids.

Hybridization protocols for the particular applications disclosed here are described in Pinkel *et al.* *Proc. Natl. Acad. Sci. USA*, 85: 9138-9142 (1988) and in EPO Pub. No. 430,402. Suitable hybridization protocols can also be found in *Methods o\in Molecular Biology Vol. 33: In Situ Hybridization Protocols*, K.H.A. Choo, ed., Humana Press, Totowa, New Jersey, (1994). In a particularly preferred embodiment, the hybridization protocol of Kallioniemi *et al., Proc. Natl Acad Sci USA,* 89: 5321-5325 (1992) is used.

Typically, it is desirable to use dual color FISH, in which two probes are utilized, each labelled by a different fluorescent dye. A test probe that hybridizes to the region of interest is labelled with one dye, and a control probe that hybridizes to a different region is labelled with a second dye. A nucleic acid that hybridizes to a stable portion of the chromosome of interest, such as the centromere region, is often most useful as the control probe. In this way, differences between efficiency of hybridization from sample to sample can be accounted for.

The FISH methods for detecting chromosomal abnormalities can be performed on nanogram quantities of the subject nucleic acids. Paraffin embedded tumor sections can be used, as can fresh or frozen material. Because FISH can be applied to the

limited material, touch preparations prepared from uncultured primary tumors can also be used (*see, e.g.,* Kallioniemi, A. *et al., Cytogenet. Cell Genet.* 60: 190-193 (1992)). For instance, small biopsy tissue samples from tumors can be used for touch preparations (*see, e.g.,* Kallioniemi, A. *et al., Cytogenet. Cell Genet.* 60: 190-193 (1992)). Small numbers of cells obtained from aspiration biopsy or cells in bodily fluids (*e.g.,* blood, urine, sputum and the like) can also be analyzed. For prenatal diagnosis, appropriate samples will include amniotic fluid and the like.

Southern Blots

In a Southern Blot, a genomic or cDNA (typically fragmented and separated on an electrophoretic gel) is hybridized to a probe specific for the target region. Comparison of the intensity of the hybridization signal from the probe for the target region (*e.g.,* 20q13) with the signal from a probe directed to a control (non amplified) such as centromeric DNA, provides an estimate of the relative copy number of the target nucleic acid.

Detecting Mutations in Genes from the 20q13 Amplicon

The cDNA sequences disclosed here can also be used for detecting mutations (*e.g.,* substitutions, insertions, and deletions) within the corresponding endogenous genes. One of skill will recognize that the nucleic acid hybridization techniques generally described above can be adapted to detect such much mutations. For instance, oligonucleotide probes that distinguish between mutant and wild-type forms of the target gene can be be used in standard hybridization assays. In some embodiments, amplification (*e.g.,* using PCR) can be used to increase copy number of the target sequence prior to hybridization.

Assays for detecting 20q13 amplicon proteins.

As indicated above, this invention identifies protein products of genes in the 20q13 amplicon that are associated with various cancers. In particular, it was shown that 20q13 proteins (*e.g.,* @@were overexpressed in various cancers. The presence or absence and/or level of expression of 20q13 proteins can be indicative of the presence, absence, or extent of a cancer. Thus, 20q13 proteins can provide useful diagnostic markers.

The 20q13 amplicon proteins (*e.g.*, ZABC1 or 1b1) can be detected and quantified by any of a number of means well known to those of skill in the art. These may include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC),

5   hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay(RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, western blotting, and the like.

In one preferred embodiment, the 20q13 amplicon proteins are detected in

10   an electrophoretic protein separation such as a one dimensional or two-dimensional electrophoresis, while in a most preferred embodiment, the 20q13 amplicon proteins are detected using an immunoassay.

As used herein, an immunoassay is an assay that utilizes an antibody to specifically bind to the analyte (*e.g.*, ZABC1 or 1b1 proteins). The immunoassay is thus

15   characterized by detection of specific binding of a 20q13 amplicon protein to an anti-20q13 amplicon antibody (*e.g.*, anti-ZABC1 or anti-1b1) as opposed to the use of other physical or chemical properties to isolate, target, and quantify the analyte.

The collection of biological sample and subsequent testing for 20q13 amplicon protein(s) is discussed in more detail below.

20

A) Sample Collection and Processing

The 20q13 amplicon proteins are preferably quantified in a biological sample derived from a mammal, more preferably from a human patient or from a porcine, murine, feline, canine, or bovine. As used herein, a biological sample is a sample of biological tissue

25   or fluid that contains a 20q13 amplicon protein concentration that may be correlated with a 20q13 amplification. Particularly preferred biological samples include, but are not limited to biological fluids such as blood or urine, or tissue samples including, but not limited to tissue biopsy (*e.g.*, needle biopsy) samples.

The biological sample may be pretreated as necessary by dilution in an

30   appropriate buffer solution or concentrated, if desired. Any of a number of standard aqueous buffer solutions, employing one of a variety of buffers, such as phosphate, Tris, or the like, at physiological pH can be used.

B) Electrophoretic Assays.

As indicated above, the presence or absence of 20q13 amplicon proteins in a biological tissue may be determined using electrophoretic methods. Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (*see generally*, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc., N.Y.). In a preferred embodiment, the 20q13 amplicon proteins are detected using one-dimensional or two-dimensional electrophoresis. A particularly preferred two-dimensional electrophoresis separation relies on isoelectric focusing (IEF) in immobilized pH gradients for one dimension and polyacrylamide gels for the second dimension. Such assays are described in the cited references and by Patton *et al.* (1990) *Biotechniques* 8: 518.

C) Immunological Binding Assays.

In a preferred embodiment, the 20q13 amplicon are detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.*, U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Asai, ed. Academic Press, Inc. New York (1993); *Basic and Clinical Immunology* 7th Edition, Stites & Terr, eds. (1991).

Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (in this case 20q13 amplicon). The capture agent is a moiety that specifically binds to the analyte. In a preferred embodiment, the capture agent is an antibody that specifically binds 20q13 amplicon protein(s).

The antibody (*e.g.*, anti-ZABC1 or anti-1b1) may be produced by any of a number of means well known to those of skill in the art (*see, e.g. Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Asai, ed. Academic Press, Inc. New York (1993); and *Basic and Clinical Immunology* 7th Edition, Stites & Terr, eds. (1991)). The antibody may be a whole antibody or an antibody fragment. It may be polyclonal or monoclonal, and it may be produced by challenging an organism (*e.g.* mouse, rat, rabbit, *etc.*) with a 20q13 amplicon protein or an epitope derived therefrom. Alternatively, the

antibody may be produced *de novo* using recombinant DNA methodology. The antibody can also be selected from a phage display library screened against 20q13 amplicon (*see, e.g.* Vaughan *et al.* (1996) *Nature Biotechnology,* 14: 309-314 and references therein).

5   Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Thus, the labeling agent may be a labeled 20q13 amplicon protein or a labeled anti-20q13 amplicon antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/20q13 amplicon protein complex.

10   In a preferred embodiment, the labeling agent is a second human 20q13 amplicon protein antibody bearing a label. Alternatively, the second 20q13 amplicon protein antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled 15 molecule can specifically bind, such as enzyme-labeled streptavidin.

       Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of 20   species. See, generally Kronval, *et al., J. Immunol.,* 111:1401-1406 (1973), and Akerstrom, *et al., J. Immunol.,* 135:2589-2542 (1985).

       Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the 25   incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

30   **1) Non-Competitive Assay Formats**.

       Immunoassays for detecting 20q13 amplicon proteins may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the

amount of captured analyte (in this case 20q13 amplicon) is directly measured. In one preferred "sandwich" assay, for example, the capture agent (anti-20q13 amplicon protein antibodies) can be bound directly to a solid substrate where they are immobilized. These immobilized antibodies then capture 20q13 amplicon protein present in the test sample.

5   The 20q13 amplicon protein thus immobilized is then bound by a labeling agent, such as a second human 20q13 amplicon protein antibody bearing a label. Alternatively, the second 20q13 amplicon protein antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which

10   a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

2. Competitive assay formats.

In competitive assays, the amount of analyte (20q13 amplicon protein) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte (20q13 amplicon proteins such as ZABC1 or 1b1 protein) displaced

15   (or competed away) from a capture agent (*e.g.*, anti-ZABC1 or anti-1b1 antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case, 20q13 amplicon protein is added to the sample and the sample is then contacted with a capture agent, in this case an antibody that specifically binds 20q13 amplicon protein. The

20   amount of 20q13 amplicon protein bound to the antibody is inversely proportional to the concentration of 20q13 amplicon protein present in the sample.

In a particularly preferred embodiment, the anti-20q13 protein antibody is immobilized on a solid substrate. The amount of 20q13 amplicon protein bound to the antibody may be determined either by measuring the amount of 20q13 amplicon present in

25   an 20q13 amplicon protein/antibody complex, or alternatively by measuring the amount of remaining uncomplexed 20q13 amplicon protein. The amount of 20q13 amplicon protein may be detected by providing a labeled 20q13 amplicon protein.

A hapten inhibition assay is another preferred competitive assay. In this assay a known analyte, in this case 20q13 amplicon protein is immobilized on a solid

30   substrate. A known amount of anti-20q13 amplicon protein antibody is added to the sample, and the sample is then contacted with the immobilized 20q13 amplicon protein. In this case, the amount of anti-20q13 amplicon protein antibody bound to the immobilized

20q13 amplicon protein is inversely proportional to the amount of 20q13 amplicon protein present in the sample. Again the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

3. Other Assay Formats

In a particularly preferred embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of 20q13 amplicon protein in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind 20q13 amplicon protein. The anti-20q13 amplicon protein antibodies specifically bind to 20q13 amplicon protein on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the anti-20q13 amplicon protein.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (see, Monroe et al. (1986) Amer. Clin. Prod. Rev. 5:34-41).

D) Reduction of Non-Specific Binding.

One of skill in the art will appreciate that it is often desirable to reduce non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

E) Labels.

The particular label or detectable group used in the assay is not a critical aspect of the invention, so long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having

5   a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads

10   (*e.g.* Dynabeads^{TM}), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads.

15   The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

20   Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (*e.g.*, biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (*e.g.*, streptavidin) molecule which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used.

25   Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with the labeled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody.

The molecules can also be conjugated directly to signal generating

30   compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein

and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, *etc.*
Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, *e.g.,*
luminol. For a review of various labeling or signal producing systems which may be used,
see, U.S. Patent No. 4,391,904).

Means of detecting labels are well known to those of skill in the art. Thus,
for example, where the label is a radioactive label, means for detection include a
scintillation counter or photographic film as in autoradiography. Where the label is a
fluorescent label, it may be detected by exciting the fluorochrome with the appropriate
wavelength of light and detecting the resulting fluorescence. The fluorescence may be
detected visually, by means of photographic film, by the use of electronic detectors such as
charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic
labels may be detected by providing the appropriate substrates for the enzyme and
detecting the resulting reaction product. Finally simple colorimetric labels may be
detected simply by observing the color associated with the label. Thus, in various dipstick
assays, conjugated gold often appears pink, while various conjugated beads appear the
color of the bead.

Some assay formats do not require the use of labeled components. For
instance, agglutination assays can be used to detect the presence of the target antibodies.
In this case, antigen-coated particles are agglutinated by samples comprising the target
antibodies. In this format, none of the components need be labeled and the presence of the
target antibody is detected by simple visual inspection.

G) Substrates.

As mentioned above, depending upon the assay, various components,
including the antigen, target antibody, or anti-human antibody, may be bound to a solid
surface. Many methods for immobilizing biomolecules to a variety of solid surfaces are
known in the art. For instance, the solid surface may be a membrane (*e.g.*,
nitrocellulose), a microtiter dish (*e.g.*, PVC, polypropylene, or polystyrene), a test tube
(glass or plastic), a dipstick (*e.g.* glass, PVC, polypropylene, polystyrene, latex, and the
like), a microcentrifuge tube, or a glass or plastic bead. The desired component may be
covalently bound or noncovalently attached through nonspecific bonding.

A wide variety of organic and inorganic polymers, both natural and synthetic may be employed as the material for the solid surface. Illustrative polymers include polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephthalate), rayon, nylon, poly(vinyl butyrate),

5   polyvinylidene difluoride (PVDF), silicones, polyformaldehyde, cellulose, cellulose acetate, nitrocellulose, and the like. Other materials which may be employed, include paper, glasses, ceramics, metals, metalloids, semiconductive materials, cements or the like. In addition, are included substances that form gels, such as proteins (*e.g.*, gelatins), lipopolysaccharides, silicates, agarose and polyacrylamides can be used. Polymers which

10   form several aqueous phases, such as dextrans, polyalkylene glycols or surfactants, such as phospholipids, long chain (12-24 carbon atoms) alkyl ammonium salts and the like are also suitable. Where the solid surface is porous, various pore sizes may be employed depending upon the nature of the system.

In preparing the surface, a plurality of different materials may be employed,

15   particularly as laminates, to obtain various properties. For example, protein coatings, such as gelatin can be used to avoid non-specific binding, simplify covalent conjugation, enhance signal detection or the like.

If covalent bonding between a compound and the surface is desired, the surface will usually be polyfunctional or be capable of being polyfunctionalized.

20   Functional groups which may be present on the surface and used for linking can include carboxylic acids, aldehydes, amino groups, cyano groups, ethylenic groups, hydroxyl groups, mercapto groups and the like. The manner of linking a wide variety of compounds to various surfaces is well known and is amply illustrated in the literature. See, for example, *Immobilized Enzymes*, Ichiro Chibata, Halsted Press, New York, 1978, and

25   Cuatrecasas (1970) *J. Biol. Chem.* 245 3059).

In addition to covalent bonding, various methods for noncovalently binding an assay component can be used. Noncovalent binding is typically nonspecific absorption of a compound to the surface. Typically, the surface is blocked with a second compound to prevent nonspecific binding of labeled assay components. Alternatively, the surface is

30   designed such that it nonspecifically binds one component but does not significantly bind another. For example, a surface bearing a lectin such as Concanavalin A will bind a carbohydrate containing compound but not a labeled protein that lacks glycosylation.

Various solid surfaces for use in noncovalent attachment of assay components are reviewed in U.S. Patent Nos. 4,447,576 and 4,254,082.

Kits Containing 20q13 Amplicon Probes.

This invention also provides diagnostic kits for the detection of chromosomal abnormalities at 20q13. In a preferred embodiment, the kits include one or more probes to the 20q13 amplicon and/or antibodies to a 20q13 amplicon (*e.g.*, anti-ZABC1 *or anti-1b1*) described herein. The kits can additionally include blocking probes, instructional materials describing how to use the kit contents in detecting 20q13 amplicons. The kits may also include one or more of the following: various labels or labeling agents to facilitate the detection of the probes, reagents for the hybridization including buffers, a metaphase spread, bovine serum albumin (BSA) and other blocking agents, sampling devices including fine needles, swabs, aspirators and the like, positive and negative hybridization controls and so forth.

Expression of cDNA clones

One may express the desired polypeptides encoded by the cDNA clones disclosed here in a recombinantly engineered cell such as bacteria, yeast, insect (especially employing baculoviral vectors), and mammalian cells. It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of the cDNAs. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of natural or synthetic nucleic acids encoding polypetides of the invention will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding the polypeptides. To obtain high level expression of a cloned gene, it is desirable to construct expression plasmids which contain,

at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator.

Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky, C., 1984, J. Bacteriol., 158:1018-1024 and the leftward promoter of phage lambda (P_L) as described by Herskowitz, I. and Hagen, D., 1980, Ann. Rev. Genet., 14:399-445. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol. Expression systems are available using *E. coli, Bacillus* sp. (Palva, I *et al.*, 1983, Gene 22:229-235; Mosbach, K. *et al.* Nature, 302:543-545 and *Salmonella. E. coli* systems are preferred.

The polypeptides produced by prokaryote cells may not necessarily fold properly. During purification from *E. coli*, the expressed polypeptides may first be denatured and then renatured. This can be accomplished by solubilizing the bacterially produced proteins in a chaotropic agent such as guanidine HCl and reducing all the cysteine residues with a reducing agent such as beta-mercaptoethanol. The polypeptides are then renatured, either by slow dialysis or by gel filtration. U.S. Patent No. 4,511,503.

A variety of eukaryotic expression systems such as yeast, insect cell lines and mammalian cells, are known to those of skill in the art. As explained briefly below, the polypeptides may also be expressed in these eukaryotic systems.

Synthesis of heterologous proteins in yeast is well known and described. *Methods in Yeast Genetics,* Sherman, F., *et al.*, Cold Spring Harbor Laboratory, (1982) is a well recognized work describing the various methods available to produce the polypeptides in yeast. A number of yeast expression plasmids like YEp6, YEp13, YEp4 can be used as vectors. A gene of interest can be fused to any of the promoters in various yeast vectors. The above-mentioned plasmids have been fully described in the literature (Botstein, *et al.*, 1979, Gene, 8:17-24; Broach, *et al.*, 1979, Gene, 8:121-133).

Illustrative of cell cultures useful for the production of the polypeptides are cells of insect or mammalian origin. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. Illustrative examples of mammalian cell lines include VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, Cos-7 or MDCK cell lines.

As indicated above, the vector, e. g., a plasmid, which is used to transform the host cell, preferably contains DNA sequences to initiate transcription and sequences to control the translation of the antigen gene sequence. These sequences are referred to as expression control sequences. When the host cell is of insect or mammalian origin

5   illustrative expression control sequences are often obtained from the SV-40 promoter (Science, 222:524-527, 1983), the CMV I.E. Promoter (Proc. Natl. Acad. Sci. 81:659-663, 1984) or the metallothionein promoter (Nature 296:39-42, 1982). The cloning vector containing the expression control sequences is cleaved using restriction enzymes and adjusted in size as necessary or desirable and ligated with the desired DNA

10   by means well known in the art.

As with yeast, when higher animal host cells are employed, polyadenlyation or transcription terminator sequences from known mammalian genes need to be incorporated into the vector. An example of a terminator sequence is the polyadenlyation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the

15   transcript may also be included. An example of a splicing sequence is the VPl intron from SV40 (Sprague, J. et al., 1983, J. Virol. 45: 773-781).

Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., 1985, "Bovine Papilloma virus DNA a Eukaryotic Cloning Vector"

20   in DNA Cloning Vol. II a Practical Approach Ed. D.M. Glover, IRL Press, Arlington, Virginia pp. 213-238.

Therapeutic and other uses of cDNAs and their gene products

The cDNA sequences and the polypeptide products of the invention can be

25   used to modulate the activity of the gene products of the endogenous genes corresponding to the cDNAs. By modulating activity of the gene products, pathological conditions associated with their expression or lack of expression can be treated. Any of a number of techniques well known to those of skill in the art can be used for this purpose.

The cDNAs of the invention are particularly used for the treatment of

30   various cancers such as cancers of the breast, ovary, bladder, head and neck, and colon. Other diseases may also be treated with the sequences of the invention. For instance, as noted above, GCAP (SEQ. ID. No. 6) encodes a guanino cyclase activating protein which

is involved in the biosynthesis of cyclic AMP. Mutations in genes involved in the biosynthesis of cyclic AMP are known to be associated with hereditary retinal degenerative diseases. These diseases are a group of inherited conditions in which progressive, bilateral degeneration of retinal structures leads to loss of retinal function. These diseases include age-related macular degeneration, a leading cause of visual impairment in the elderly; Leber's congenital amaurosis, which causes its victims to be born blind; and retinitis pigmentosa ("RP"), one of the most common forms of inherited blindness. RP is the name given to those inherited retinopathies which are characterized by loss of retinal photoreceptors (rods and cones), with retinal electrical responses to light flashes (i.e. eletroretinograms, or "ERGs") that are reduced in amplitude.

The mechanism of retinal photoreceptor loss or cell death in different retinal degenerations is not fully understood. Mutations in a number of different genes have been identified as the primary genetic lesion in different forms of human RP. Affected genes include rhodopsin, the alpha and beta subunits of cGMP photodiesterase, and peripherin-RDS (Dryja, T. P. et al., *Invest. Ophthalmol. Vis. Sci.* 36, 1197-1200 (1995)). In all cases the manifestations of the disorder regardless of the specific primary genetic mutation is similar, resulting in photoreceptor cell degeneration and blindness.

Studies on animal models of retinal degeneration have been the focus of many laboratories during the last decade. The mechanisms that are altered in some of the mutations leading to blindness have been elucidated. This would include the inherited disorders of the *rd* mouse. The *rd* gene encodes the beta subunit of cGMP-phosphodiesterase (PDE) (Bowes, C. et al., *Nature* 347, 677-680 (1990)), an enzyme of fundamental importance in normal visual function because it is a key component in the cascade of events that takes place in phototransduction.

The polypeptides encoded by the cDNAs of the invention can be used as immunogens to raise antibodies either polyclonal or monoclonal. The antibodies can be used to detect the polypeptides for diagnostic purposes, as therapeutic agents to inhibit the polypeptides, or as targeting moieties in immunotoxins. The production of monoclonal antibodies against a desired antigen is well known to those of skill in the art and is not reviewed in detail here.

Those skilled in the art recognize that there are many methods for production and manipulation of various immunoglobulin molecules. As used herein, the

terms "immunoglobulin" and "antibody" refer to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. Immunoglobulins may exist in a variety of forms besides antibodies, including for example, Fv, Fab, and F(ab)$_2$, as well as in single chains. To raise monoclonal antibodies, antibody-producing cells obtained from immunized animals (e.g., mice) are immortalized and screened, or screened first for the production of the desired antibody and then immortalized. For a discussion of general procedures of monoclonal antibody production see Harlow and Lane, *Antibodies, A Laboratory Manual* Cold Spring Harbor Publications, N.Y. (1988).

The antibodies raised by these techniques can be used in immunodiagnostic assays to detect or quantify the expression of gene products from the nucleic acids disclosed here. For instance, labeled monoclonal antibodies to polypeptides of the invention can be used to detect expression levels in a biological sample. For a review of the general procedures in diagnostic immunoassays, see *Basic and Clinical Immunology* 7th Edition D. Stites and A. Terr ed. (1991).

The polynucleotides of the invention are particularly useful for gene therapy techniques well known to those skilled in the art. Gene therapy as used herein refers to the multitude of techniques by which gene expression may be altered in cells. Such methods include, for instance, introduction of DNA encoding ribozymes or antisense nucleic acids to inhibit expression as well as introduction of functional wild-type genes to replace mutant genes (e.g., using wild-type GCAP genes to treat retinal degeneration). A number of suitable viral vectors are known. Such vectors include retroviral vectors (see Miller, *Curr. Top. Microbiol. Immunol.* 158: 1-24 (1992); Salmons and Gunzburg, *Human Gene Therapy* 4: 129-141 (1993); Miller et al., *Methods in Enzymology* 217: 581-599, (1994)) and adeno-associated vectors (reviewed in Carter, *Curr. Opinion Biotech.* 3: 533-539 (1992); Muzcyzka, *Curr. Top. Microbiol. Immunol.* 158: 97-129 (1992)). Other viral vectors that may be used within the methods include adenoviral vectors, herpes viral vectors and Sindbis viral vectors, as generally described in, e.g., Jolly, *Cancer Gene Therapy* 1:51-64 (1994); Latchman, *Molec. Biotechnol.* 2:179-195 (1994); and Johanning et al., *Nucl. Acids Res.* 23:1495-1501 (1995).

Delivery of nucleic acids linked to a heterologous promoter-enhancer element via liposomes is also known (see, e.g., Brigham, et al. (1989) *Am. J. Med. Sci.*, 298:278-281; Nabel, et al. (1990) *Science*, 249:1285-1288; Hazinski, et al. (1991) *Am. J.*

Resp. Cell Molec. Biol., 4:206-209; and Wang and Huang (1987) *Proc. Natl. Acad. Sci. (USA)*, 84:7851-7855); coupled to ligand-specific, cation-based transport systems (Wu and Wu (1988) *J. Biol. Chem.*, 263:14621-14624). Naked DNA expression vectors have also been described (Nabel *et al.* (1990), *supra*); Wolff *et al.* (1990) *Science*, 247:1465-1468).

5   The nucleic acids and encoded polypeptides of the invention can be used directedly to inhibit the endogenous genes or their gene products. For instance, Inhibitory nucleic acids may be used to specifically bind to a complementary nucleic acid sequence. By binding to the appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex is formed. These nucleic acids are often termed "antisense" because

10   they are usually complementary to the sense or coding strand of the gene, although approaches for use of "sense" nucleic acids have also been developed. The term "inhibitory nucleic acids" as used herein, refers to both "sense" and "antisense" nucleic acids. Inhibitory nucleic acid methods encompass a number of different approaches to altering expression of specific genes that operate by different mechanisms.   In

15   brief, inhibitory nucleic acid therapy approaches can be classified into those that target DNA sequences, those that target RNA sequences (including pre-mRNA and mRNA), those that target proteins (sense strand approaches), and those that cause cleavage or chemical modification of the target nucleic acids (ribozymes). These different types of inhibitory nucleic acid technology are described, for instance, in Helene, C. and Toulme,

20   J. (1990) *Biochim. Biophys. Acta.*, 1049:99-125. Inhibitory nucleic acid complementary to regions of c-myc mRNA has been shown to inhibit c-myc protein expression in a human promyelocytic leukemia cell line, HL60, which overexpresses the c-myc protoncogene. *See* Wickstrom E.L., *et al.*, (1988) *PNAS (USA)*, 85:1028-1032 and Harel-Bellan, A., *et al.*, (1988) *Exp. Med.*, 168:2309-2318.

25   The encoded polypeptides of the invention can also be used to design molecules (peptidic or nonpeptidic) that inhibit the endogenous proteins by, for instance, inhibiting interaction between the protein and a second molecule specifically recognized by the protein. Methods for designing such molecules are well known to those skilled in the art.

30   For instance, polypeptides can be designed which have sequence identity with the encoded proteins or may comprise modifications (conservative or non-conservative) of the sequences. The modifications can be selected, for example, to alter

their *in vivo* stability. For instance, inclusion of one or more D-amino acids in the peptide typically increases stability, particularly if the D-amino acid residues are substituted at one or both termini of the peptide sequence.

The polypeptides can also be modified by linkage to other molecules. For example, different N- or C-terminal groups may be introduced to alter the molecule's physical and/or chemical properties. Such alterations may be utilized to affect, for example, adhesion, stability, bio-availability, localization or detection of the molecules.

For diagnostic purposes, a wide variety of labels may be linked to the terminus, which may provide, directly or indirectly, a detectable signal. Thus, the polypeptides may be modified in a variety of ways for a variety of end purposes while still retaining biological activity.

EXAMPLES

The following examples are offered to illustrate, but not to limit the present invention.

Example 1

PROGNOSTIC IMPLICATIONS OF AMPLIFICATION OF CHROMOSOMAL REGION 20q13 IN BREAST CANCER

Patients and tumor material.

Tumor samples were obtained from 152 women who underwent surgery for breast cancer between 1987 and 1992 at the Tampere University or City Hospitals. One hundred and forty-two samples were from primary breast carcinomas and 11 from metastatic tumors. Specimens from both the primary tumor and a local metastasis were available from one patient. Ten of the primary tumors that were either *in situ* or mucinous carcinomas were excluded from the material, since the specimens were considered inadequate for FISH studies. Of the remaining 132 primary tumors, 128 were invasion ductal and 4 lobular carcinomas. The age of the patients ranged from 29 to 92 years (mean 61). Clinical follow-up was available from 129 patients. Median follow-up period was 45 months (range 1.4-1.77 months). Radiation therapy was given to 77 of the 129 patients (51 patients with positive and 26 with negative lymph nodes), and systemic adjuvant therapy to 36 patients (33 with endocrine and 3 with cytotoxic chemotherapy). Primary tumor size and axillary node involvement were determined according to the tumor-node

metastasis (TNM) classification. The histopathological diagnosis was evaluated according to the World Health Organization (11). The carcinomas were graded on the basis of the tubular arrangement of cancer cells, nuclear atypia, and frequency of mitotic or hyperchromatic nuclear figures according to Bloom and Richardson, *Br. J. Cancer*, 11: 359-377 (1957).

Surgical biopsy specimens were frozen at -70°C within 15 minutes of removal. Cryostat sections (5-6 μm) were prepared for intraoperative histopathological diagnosis, and additional thin sections were cut for immunohistochemical studies. One adjacent 200 μm thick section was cut for DNA flow cytometric and FISH studies.

Cell preparation for FISH.

After histological verification that the biopsy specimens contained a high proportion of tumor cells, nuclei were isolated from 200 μm frozen sections according to a modified Vindelov procedure for DNA flow cytometry, fixed and dropped on slides for FISH analysis as described by Hyytinen *et al., Cytometry* 16: 93-99 (1994). Foreskin fibroblasts were used as negative controls in amplification studies and were prepared by harvesting cells at confluency to obtain G1 phase enriched interphase nuclei. All samples were fixed in methanol-acetic-acid (3:1).

Probes.

Five probes mapping to the 20q13 region were used (*see* Stokke, *et al., Genomics*, 26: 134-137 (1995)). The probes included Pl-clones for melanocortin-3-receptor (probe MC3R, fractional length from p-arm telomere (Flpter 0.81) and phosphoenolpyruvate carboxy kinase (PCK, Flpter 0.84), as well as anonymous cosmid clones RMC20C026 (Flpter 0.79). In addition, RMC20C001 (Flpter 0.825) and RMC20C030 (Flpter 0.85) were used. Probe RMC20C001 was previously shown to define the region of maximum amplification (Tanner *et al., Cancer Res*, 54: 4257-4260 (1994)). One cosmid probe mapping to the proximal p-arm, RMC20C038 (FLpter 0.237) was used as a chromosome-specific reference probe. Test probes were labeled with biotin-14-dATP and the reference probe with digoxigenin-11-dUTP using nick translation (Kallioniemi *et al., Proc. Natl Acad Sci USA*, 89: 5321-5325 (1992)).

Fluorescence in situ hybridization.

Two-color FISH was performed using biotin-labeled 20q13-specific probes and digoxigenin-labelled 20p reference probe essentially as described (*Id.*). Tumor samples were postfixed in 4% paraformaldtheyde/phosphate-buffered saline for 5 min at 4

5   C prior to hybridization, dehydrated in 70%, 85% and 100% ethanol, air dried, and incubated for 30 min at 80°C. Slides were denatured in a 70% formamide/2x standard saline citrate solution at 72-74°C for 3 min, followed by a proteinase K digestion (0.5 μg/ml). . The hybridization mixture contained 18 ng of each of the labeled probes and 10 μg human placental DNA. After hybridization, the probes were detected

10   immunochemically with avidin-FITC and anti-digoxigenin Rhodamine. Slides were counterstained with 0.2 μM 4,6-diamidino-2-phenylindole (DAPI) in an antifade solution.

Fluorescence microscopy and scoring of signals in interphase nuclei.

A Nikon fluorescence microscope equipped with double band-bass filters (Chromatechnology, Brattleboro, Vermont, USA) and 63 x objective (NA 1.3) was used

15   for simultaneous visualization of FITC and Rhodamine signals. At least 50 non-overlapping nuclei with intact morphology based on the DAPI counterstaining were scored to determine the number of test and reference probe hybridization signals. Leukocytes infiltrating the tumor were excluded from analysis. Control hybridizations to

20   normal fibroblast interphase nuclei were done to ascertain that the probes recognized a single copy target and that the hybridization efficiencies of the test and reference probes were similar.

The scoring results were expressed both as the mean number of hybridization signals per cell and as mean level of amplification (= mean of number of

25   signals relative to the number of reference probe signals).

DNA flow cytometry and steroid receptor analyses.

DNA flow cytometry was performed from frozen 200 μm sections as described by Kallioniemi, *Cytometry* 9: 164-169 (1988). Analysis was carried out using an

30   EPICS C flow cytometer (Coulter Electronics Inc., Hialeah, Forida, USA) and the MultiCycle program (Phoenix Flow Systems, San Diego, California, USA). DNA-index over 1.07 (in over 20% of cells) was used as a criterion for DNA aneuploidy. In DNA

aneuploid histograms, the S-phase was analyzed only from the aneuploid clone. Cell cycle evaluation was successful in 86% (108/126) of the tumors.

Estrogen (ER) and progesterone (PR) receptors were detected immunohistochemically. from cryostat sections as previously described (17). The staining

5 results were semiquantitatively evaluated and a histoscore greater than or equal to 100 was considered positive for both ER and PR (17).

Statistical Methods.

Contingency tables were analyzed with Chi square test for trend.

10 Association between S-phase fraction (continuous variable) and 20q13 amplification was analyzed with Kruskal-Wallis test. Analysis of disease-free survival was performed using the BMDPIL program and Mautel-Cox test and Cox's proportional hazards model (BMDP2L program) was used in multivariate regression analysis (Dixon *BMDP Statistical Software*. London, Berkeley, Los Angeles: University of California Press, (1981)).

15

Amplification of 20q13 in primary breast carcinomas by fluorescence *in situ* hybridization.

The minimal region probe RMC20C001 was used in FISH analysis to assess the 20q13 amplification. FISH was used to analyze both the total number of signals in

20 individual tumor cells and to determine the mean level of amplification (mean copy number with the RMC20C001 probe relative to a 20p-reference probe). In addition, the distribution of the number of signals in the tumor nuclei was also assessed. Tumors were classified into three categories: no. low and high level of amplification. Tumors classified as not amplified showed less than 1.5 than 1.5 fold-copy number of the RMC20C001 as

25 compared to the p-arm control. Those classified as having low-level amplification had 1.5-3-fold average level of amplification. Tumors showing over 3-fold average level of amplification were classified as highly amplified.

The highly amplified tumors often showed extensive intratumor heterogeneity with up to 40 signals in individual tumor cells. In highly amplified tumors,

30 the RMC20C001 probe signals were always arranged in clusters by FISH, which indicates location of the amplified DNA sequences in close proximity to one another *e.g.* in a tandem array. Low level 20q13 amplification was found in 29 of the 132 primary tumors

(22%), whereas nine cases (6.8%) showed high level amplification. The overall prevalence of increased copy number in 20q13 was thus 29% (38/132).

Defining the minimal region of amplification.

The average copy number of four probes flanking RMC20C001 was determined in the nine highly amplified tumors. The flanking probes tested were malanocortin-3-receptor (MC3R, FLpter 0.81), phosphoenolpyruvate carboxykinase (PCK, 0.84), RMC20C026 (0.79) and RMC20C030 (0.85). The amplicon size and location varied slightly from one tumor to another but RMC20C001 was the only probe consistently highly amplified in all nine cases.

Association of 20q13 amplification with pathological and biological features.

The 20q13 amplification was significantly associated with high histologic grade of the tumors (p=0.01). This correlation was seen both in moderately and highly amplified tumors (Table 4). Amplification of 20q13 was also significantly associated with aneuploidy as determined by DNA flow cytometry (p=0.01, Table 4) The mean cell proliferation activity, measured as the percentage of cells in the S-phase fraction, increased (p=0.0085 by Kruskal-Wallis test) with the level of amplification in tumors with no, low and high levels of amplification (Table 4). No association was found with the age of the patient, primary tumor size, axillary nodal or steroid hormone-receptor status (Table 4).

Table 4. Clinicopathological correlations of amplification at chromosomal region 20q13 in 132 primary breast cancers.

Pathobiologic feature	20q13 amplification status			p-value[1]
	NO Number of patients (%)	LOW LEVEL Number of patients (%)	HIGH LEVEL Number of patients (%)	
All primary tumors	94 (71%)	29 (22%)	9 (6.8%)	
Age of patients				
< 50 years	17 (65%)	6 (23%)	3 (12%)	
≥ 50 years	77 (73%)	23 (22%)	6 (5.7%)	.39
Tumor size				
< 2 cm	33 (79%)	7 (17%)	2 (4.8%)	
≥ 2 cm	58 (67%)	22 (25%)	7 (8.0%)	.16
Nodal status				
Negative	49 (67%)	19 (26%)	5 (6.8%)	
Positive	41 (75%)	10 (18%)	4 (7.3%)	.41
Histologic grade				
I - II	72 (76%)	18 (19%)	5 (5.3%)	
III	16 (52%)	11 (35%)	4 (13%)	.01
Estrogen receptor status				
Negative	30 (67%)	10 (22%)	5 (11%)	
Positive	59 (72%)	19 (23%)	4 (4.9%)	.42
Progesterone receptor status				
Negative	57 (69%)	20 (24%)	6 (7.2%)	
Positive	32 (74%)	8 (19%)	3 (7.0%)	.53
DNA ploidy				
Diploid	45 (82%)	8 (14.5%)	2 (3.6%)	.01
Aneuploid	44 (62%)	20 (28%)	7 (10%)	
S-phase fraction (%)	mean + SD 9.9 ± 7.2	mean + SD 12.6 ± 6.7	mean + SD 19.0 ± 10.5	.0085[1]

[1] Kruskal-Wallis Test.

Relationship between 20q13 amplification and disease-free survival.

Disease-free survival of patients with high-level 20q13 amplification was significantly shorter than for patients with no or only low-level amplification (p-0.04). Disease-free survival of patients with moderately amplified tumors did not differ significantly from that of patients with no amplification. Among the node-negative patients (n=79), high level 20q13 amplification was a highly significant prognostic factor for

shorter disease-free survival (p=0.002), even in multivariate Cox's regression analysis (p=0.026) after adjustment for tumor size ER, PR grade, ploidy and S-phase fraction.

20q13 amplification in metastatic breast tumors.

5   Two of 11 metastatic breast tumors had low level and one high level 20q13 amplification. Thus, the overall prevalence (27%) of increased 20q13 copy number in metastatic tumors was a similar to that observed in the primary tumors. Both a primary and a metastatic tumor specimens were available from one of the patients. This 29-year old patient developed a pectoral muscle infiltrating metastasis eight months after total

10   mastectomy. The patient did not receive adjuvant or radiation therapy after mastectomy. The majority of tumor cells in the primary tumor showed a low level amplification, although individual tumor cells (less than 5% of total) contained 8-20 copies per cell by FISH. In contrast, all tumor cells from metastasis showed high level 20q13 amplification (12-50 copies per cell). The absolute copy number of the reference probe remained the

15   same suggesting that high level amplification was not a result of an increased degree of aneuploidy.

Diagnostic and Prognostic Value of the 20q13 Amplification.

      The present findings suggest that the newly-discovered 20q13 amplification may be an important component of the genetic progression pathway of certain breast

20   carcinomas. Specifically, the foregoing experiments establish that: 1) High-level 20q13 amplification, detected in 7% of the tumors, was significantly associated with decreased disease-free survival in node-negative breast cancer patients, as well as with indirect indicators of high-malignant potential, such as high grade and S-phase fraction. 2)

25   Low-level amplification, which was much more common, was also associated with clinicopathological features of aggressive tumors, but was not prognostically significant. 3) The level of amplification of RMC20C001 remains higher than amplification of nearby candidate genes and loci indicating that a novel oncogene is located in the vicinity of RMC20C001.

30   High-level 20q13 amplification was defined by the presence of more than 3-fold higher copy number of the 20q13 amplification is somewhat lower than the amplification frequencies reported for some of the other breast cancer oncogenes, such as

ERBB2 (17q12) and Cyclin-D (11q13) (Borg *et al., Oncogene*, 6: 137-143 (1991), Van de

Vijver *et al. Adv. Canc. Res.*, 61: 25-56 (1993)). However, similar to what has been

previously found with these other oncogenes (Swab, *et al., Genes Chrom. Canc.*, 1:

181-193 (1990), Borg *et al., supra.*), high-level 20q13 amplification was more common in

5   tumors with high grade or high S-phase fraction and in cases with poor prognosis.

Although only a small number of node-negative patients was analyzed, our results suggest

that 20q13 amplification might have independent role as a prognostic indicator. Studies to

address this question in large patient materials are warranted. Moreover, based on these

survival correlations, the currently unknown, putative oncogene amplified in this locus

10   may confer an aggressive phenotype. Thus, cloning of this gene is an important goal.

Based on the association of amplification with highly proliferative tumors one could

hypothesize a role for this gene in the growth regulation of the cell.

The role of the low-level 20q13 amplification as a significant event in tumor

progression appears less clear. Low-level amplification was defined as 1.5-3-fold

15   increased average copy number of the 20q13 probe relative to the p-arm control. In

addition, these tumors characteristically lacked individual tumor cells with very high copy

numbers, and showed a scattered, not clustered, appearance of the signals. Accurate

distinction between high and low level 20q13 amplification can only be reliably done by

FISH, whereas Southern and slot blot analyses are likely to be able to detect only

20   high-level amplification, in which substantial elevation of the average gene copy number

takes place. This distinction is important, because only the high amplified tumors were

associated with adverse clinical outcome. Tumors with low-level 20q13 amplification

appeared to have many clinicopathological features that were in between of those found for

tumors with no and those with high level amplification. For example, the average tumor

25   S-phase fraction was lowest in the non-amplified tumors and highest in the highly

amplified tumors. One possibility is that low-level amplification precedes the development

of high level amplification. This has been shown to be the case, *e.g.*, in the development

of drug resistance-gene amplification *in vitro* (Stark, *Adv. Canc. Res.*, 61: 87-113 (1993)).

Evidence supporting this hypothesis was found in one of our patients, whose local

30   metastasis contained a much higher level of 20q13 amplification than the primary tumor

operated 8 months earlier.

Finally, our previous paper reported a 1.5 Mb critical region defined by RMC20C001 probe and exclusion of candidate genes in breast cancer cell lines and in a limited number of primary breast tumors. Results of the present study confirm these findings by showing conclusively in a larger set of primary tumors that the critical region

5 of amplification is indeed defined by this probe.

The present data thus suggest that the high-level 20q13 amplification may be a significant step in the progression of certain breast tumors to a more malignant phenotype. The clinical and prognostic implications of 20q13 amplification are striking and location of the minimal region of amplification at 20q13 has now been defined.

10

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent

15 applications cited herein are hereby incorporated by reference for all purposes.

SEQUENCE LISTING

SEQ. ID. No. 1

3bf4   3000 bp

CCGCCGGCCGGGGCGCCTGGCTGCACTCAGCGCCGGAGCCGGGAGCTAGCGGCCGCCGCCATGTCCCACCAGACCGGCATCC
AAGCAAGTGAAGATGTTAAAGAGATCTTTGCCAGAGCCAGAAATGGAAAGTACAGACTTCTGAAAATATCTATTGAAAATGA
GCAACTTGTGATTGGATCATATATTAGTCAGCCTTCAGATTCCTGGGATAAGGATTATGATTCCTTTGTTTTACCCCTGTTGGAG
GACAAACAACCATGCTATATATATTATTCAGGTTAGATTCTCAGAATGCCCAGGGATATGAATGGATATTCATTGCATGGTCTC
CAGATCATTCTCATGTTCGTCAAAAAATGTTGTATGCAGCAACAAGAGCAACTCTGAAGAAGGAATTTGGAGGTGGCCACAT
TAAAGATGAAGTATTTGGAACAGTAAAGGAAGATGTATCATTACATGGATATAAAAAATACTTGCTGTCACAATCTTCCCCT
GCCCCACTGACTGCAGCTGAGGAAGAACTACGACAGATTAAAATCAATGAGGTACAGACTGACGTGGGTGTGGACACTAAGC
ATCAAACACTACAAGGAGTAGCATTTCCCATTTCTCGAGAAGCCTTTCAGGCTTTGGAAAAATTGAATAATAGACAGCTCAA
CTATGTGCAGTTGGAAATAGATATAAAAAATGAAATTATAATTTTGGCCAACACAACAAATACAGAACTGAAAGATTTGCCA
AAGAGGATTCCCAAGGATTCAGCTCGTTACCATTTCTTTCTGTATAAACATTCCCATGAAGGAGACTATTTAGAGTCCATAG
TTTTTATTTATTCAATGCCTGGATACACATGCAGTATAAGAGAGCGGATGCTGTATTCTAGCTGCAAGAGCCGTCTGCTAGA
AATTGTAGAAAGACAACTACAAATGGATGTAATTAGAAAGATCGAGATAGACAATGGGGATGAGTTGACTGCAGACTTCCTT
TATGAAGAAGTACATCCCAAGCAGCATGCACACAAGCAAAGTTTTGCAAAACCAAAAGGTCCTGCAGGAAAAAGAGGAATTC
GAAGACTAATTAGGGGCCCAGCGGAAACTGAAGCTACTACTGATTAAAGTCATCACATTAAACATTGTAATACTAGTTTTTT
AAAAGTCCAGCTTTTAGTACAGGAGAACTGAAATCATTCCATGTTGATATAAAGTAGGGAAAAAAATTGTACTTTTTGGAAA
ATAGCACTTTTCACTTCTGTGTGTTTTTAAAATTAATGTTATAGAAGACTCATGATTTCTATTTTTGAGTTAAAGCTAGAAA
AGGGTTCAACATAATGTTTAATTTTGTCACACTGTTTTCATAGCGTTGATTCCACACTTCAAATACTTCTTAAAATTTTATA
CAGTTGGGCCAGTTCTAGAAAAGTCTGATGTCTCAAAGGGTAAACTTACTACTTTCTTGTGTGGGACAGAAAGACCTTAAAATAT
TCATATTACTTAATGAATATGTTAAGGACCAGGCTAGAGTATTTTCTAAGCTGGAAACTTAGTGTGCCTTGGAAAAGCCGCA
AGTTGCTTACTCCGAGTAGCTGTGCTAGCTCTGTCAGACTGTAGGATCATGTCTGCAACTTTTAGAAATAGTGCTTTATATT
GCAGCAGTCTTTTATATTTGACTTTTTTTTAATAGCATTAAAATTGCAGATCAGCTCACTCTGAAACTTTAAGGGTACCAGA
TATTTTCTATACTGCAGGATTTCTGATGACATTGAAAGACTTTAAACAGCCTTAGTAAATTATCTTTCTAATGCTCTGTGAG
GCCAAACATTTATGTTCAGATTGAAATTTAAATTAATATCATTCAAAAGGAAACAAAAAATGTTGAGTTTTAAAAATCAGGA
TTGACTTTTTTCTCCAAAACCATACATTTATGGGCAAATTGTGTTCTTTATCACTTCCGAGCAAATACTCAGATTTAAAATT
ACTTTAAAGTCCTGGTACTTAACAGGCTAACGTAGATAAACACCTTAATAATCTCAGTTAATACTGTATTTCAAAACACATT
TAACTGTTTTCTAATGCTTTGCATTATCAGTTACAACCTAGAGAGATTTTGAGCCTCATATTTCTTTGATACTTGAAATAGA
GGGAGCTAGAACACTTAATGTTTAATCTGTTAAACCTGCTGCAAGAGCCATAACTTTGAGGCATTTTCTAAATGAACTGTGG
GGATCCAGGATTTGTAATTTCTTGATCTAAACTTTATGCTGCATAAATCACTTATCGGAAATGCACATTTCATAGTGTGAAG
CACTCATTTCTAAACCTTATTATCTAAGGTAATATATGCACCTTTCAGAAATTTGTGTTCGAGTAAGTAAAGCATATTAGAA
TAATTGTGGGTTGACAGATTTTTAAAATAGAATTTAGAGTATTTGGGGTTTTGTTTGTTTACAAATAATCAGACTATAATAT
TTAAACATGCAAAATAACTGACAATAATGTTGCACTTGTTTTACTAAAGATATAAGTTGTTCCATGGGTGTACACGTAGACAG
ACACACATACACCCAAATTATTGCATTAAGAATCCTGGAGCAGACCATAGCTGAAGCTGTTATTTTCAGTCAGGAAGACTAC
CTGTCATGAAGGTATAAAAATAATTTAGAAGTGAATGTTTTTCTGTACCATCTATGTGCAATTATACTCTAAATTCCACTACA
CTACATTAAAGTAAATGGACATTCCAGAATATAGATGTGATTATAGTCTTAAACTAATTATTATTAAACCAATGATTGCTGA
AAATCAGTGATGCATTTGTTATAGAGTATAACTCATCGTTTACAGTATGTTTTAGTTGGCAGTATCATACCTAGATGGTGAA
TAACATATTCCCAGTAAATTTATATAGCAGTGAAGAATTACATGCCTTCTGGTGGACATTTTATAAGTGCATTTTATATCAC
AATAAAAATTTTTTCTCTTTAAAAAAAAAAAAAACAAGAAAAAAAAAAAA

SEQ. ID. No. 2

1b11   723 bp

TGGAAGCTGTCATGGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATTATTCCTGGTTTCTAGACAGGCCAAATGTAA
TTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAGATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAA
AGGACCTCTTATCCGTTCTTCCCCTTGGGGAAGGGCTTTTGCCACTTCCATGTCAATGTGGCAGTTGAGCTTGGAAATTGGT
GCGTTGTACAACATAAGCATTACTTCTCCAAGATGTGCCTGTGTAGAAATGGTCATAGATTCAAAACTGTAGCTACTATGTG
GACAGGGGGGCAGCAAGGACCCCACTTTGTAAAACATGTTTTGGGGGAATGTTTTGTTTTTCATTTCTTATTACCTGGCAA
AATAATCCAGGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTC
AACGAAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAAGAGAAGAAAGCAAACCAAGAACCTCTCTGATGRCGT
TTCTCAGACAAATGGTAAGCCCCTTACTTCCAGTATAGGAAACCTAAGATACCTAGAGCGGCTTTTGGGAACAATGGGCTCA
TGCCACAGGTAGTAGGAGACATAATTGTAGCTGGTGTGTATGGAATGTGAATGGAATATGGATTGCG

SEQ. ID. No. 3

cc49   1507 bp

GCAGGTTGCTGGGATTGACTTCTTGCTCAATTGAAACACTCATTCAATGGAGACAAAGAGCACTAATGCTTTGTGCTGATTC
ATATTTGAATCGAGGCATTGGGAACCCTGTATGCCTTGTTTGTGGAAAGAACCAGTGACACCATCACTGAGCTTCCTAAAAG
TTCGAAGAAGTTAGAGGACTATACACTTTCTTTTGAACTTTTATAATAAATATTTGCTCTGGTTTTGGAACCCAGGACTGTT
AGAGGGTGAGTGACAGGTCTTACAGTGGCCTTAATCCAACTCCAGAAATTGCCCAACGGAACTTTGAGATTATATGCAATCG
AAAGTGACAGGAAACATGCCAACTCAATCCCTCTTAATGTACATGGATGGCCAAGAGTGATTGGCAGCTCTCTTGCCAGTCC
GATGGAGATGGAGATGCCTTGTCAATGAAAGGGCCCNCTGTTGTCAATTCCGAGCTACACAAAGAAAAAAATGTCAATCCGA
ATCGAGGGGAATATGCCCTTGGATTGCATGTTCTGCAGCCAGACCTTCACACATTCAGAAGACCTTAATAAACATGTCTTAA
TGCAACACCGGCCTACCCTCTGTGAACCAGCAGTTCTTCGGGTTGAAGCAGAGTATCTCAGTCCGCTTGATAAAAGTCAAGT

GCGAACAGAACCTCCCAAGGAAAAGAATTGCAAGGAAAATGAATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTCGCT
TTTGATGTTGAGATCCACATGAGAACACACAAAGATTCTTTCACTTACGGGTGTAACATGTGCGGAAGAAGATTCAAGGAGC
CTTGGTTTCTTAAAAATCACATGCGGACRCATAATGGCAAATCGGGGGCCAGAAGCAAACTGCAGCAAGGCTTGGAGAGTAG
TCCAGCAACGATCAACGAGGTCGTCCAGGTGCACGCGGCCGAGAGCATCTCCTCTCCTTGCAAAATCTGCATGGTTTGTGGC
TTCCTATTTCCAAATAAAGAAAGTCTAATTGAGCACCGCAAGGTGCACACCAAAAAAACTGCTTTCGGTACCAGCAGCGCGC
AGACAGACTCTCCACAAGGAGGAATGCCGTCCTCGAGGGAGGACTTCCTGCAGTTGTTCAACTTGAGACCAAAATCTCACCC
TGAAACGGGGAAGAAGCCTGTCAGATGCATCCCTCAGCTCGATCCGTTCACCACCTTCCAGGCTTGGCAKCTGGCTACCAAA
GGAAWAGTTGCCATTTGCCAAGAAGTGAAGGAATTGGGGCAAGAAGGGAGCACCGACAACGACGATTCGAGTTCCGAGAAGG
AGCTTGGAGAAACAAATAAGAACCATTGTGCAGGCCTCTCGCAAGAGAAAGAGAAGTGCAAACACTCCCACGGCGAAGCGCC
CTCCGTGGACGCGGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTCACTGCTCCGAGTGCGGCAAAGCTTTCAGAACC
TACCACCAGCTGGTCTTGCACTCCAGGGTCC

SEQ. ID. No. 4
cc43   2605 bp
CAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCC
CCCGGGCTGCAGGAATTCGGCACGAGCTGGGCTACTACGATGGCGATGAGTTTCGAGTGGCCGTGGCAGTATCGCTTCCCAC
CCTTCTTTACGTTACAACCGAATGTGGACACTCGGCAGAAGCAGCTGGCCGCCTGGTGCTCGCTGGTCCTGTCCTTCTGCCG
CCTGCACAAACAGTCCAGCATGACGGTGATGGAAGCTCAGGAGAGCCCGCTCTTCAACAACGTCAAGCTACAGCGAAAGCTT
CCTGTGGAGTCGATCCAGATTGTATTAGAGGAACTGAGGAAGAAAGGGAACCTCGAGTGGTTGGATAAGAGCAAGTCCAGCT
TCCTGATCATGTGGCGGAGGCCAGAAGAATGGGGGAAACTCATCTATCAGTGGGTTTCCAGGAGTGGCCAGAACAACTCCGT
CTTTACCCTGTATGAACTGACTAATGGGGAAGACACAGAGGATGAGGAGTTCCACGGGCTGGATGAAGCCACTCTACTGCGG
GCTCTGCAGGCCCTACAGCAGGAGCACAAGGCCGAGATCATCACTGTCAGCGATGGCCGAGGCGTCAAGTTCTTCTAGCAGG
GACCTGTCTCCCTTTACTTCTTACCTCCCACCTTTCCAGGGCTTTCAAAAGGAGACAGACCCAGTGTCCCCCAAAGACTGGA
TCTGTGACTCCACCAGACTCAAAAGGACTCCAGTCCTGAAGGCTGGGACCTGGGGATGGGTTTTCTCACACCCCATATGTCTG
TCCCTTGGATAGGGTGAGGCTGAAGCACCAGGGAGAAAATATGTGCTTCTTCTCGCCCTACCTCCTTTCCCATCCTAGACTG
TCCTTGAGCCAGGGTCTGTAAACCTGACACTTTATATGTGTTCACACATGTAAGTACATACACACATGCGCCTGCAGCACAT
GCTTCTGTCTCCTCCTCCTCCCACCCCTTTAGCTGCTGTTGCCTCCCTTCTCAGGCTGGTGCTGGATCCTTCCTAGGGGATG
GGGGAAGCCCTGGCTGCAGGCAGCCTTCCAGGCAATATGAAGATAGGAGGCCCACGGGCCTGGCAGTGAGAGGTGTGGCCCC
ACACCGATTTATGATATTAAAATCTCAACTCCCAAAAAAAAAAAAAAAAAAAACTGAGACTAGTTCTCTCTCTCGAGAACT
AGTCTCGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCTTTAAGGATTTATTTATTGTTTCCTCTTTA
CAGTGTCCACTTTTCTCTACTTAATACTACTTTCCAGTCTCAGAAGCCCAGAGGGAAAAAAAAAAGACCATGAATCTTCCTC
TCCCAGATTAAAGTACACACTTTGGAAAACAGATTGGAAAACCTTTCTGAAAAAAGTTGACTGAAACTCCAAACCAACATGC
CATATTGTTGATGTTGCTCATGAAAATTGTTAAAAAACCTGTTCTAGATAAAGAACAGTCTCAAGTTTTTGTACAGCCTACAC
ATAGTACAAGGGTCCCCTATGATGATTCTTCTGTAGGACGAAATAATGTAATTTTTTCAGTTTCTGGTTTATAACTCTCTCG
ATCTCAGAGTTGACTGATTAAAACACCTACTCATGCAACAGAGAATAAAGCACTCATATTTTTATAAATTATATGGACCAAA
CTATTTTGGAAATCTTATCTATTGGAGACACAATATGCTGGACTAAAGCAATAATTATTTTATTCTCAATGTCTGTGCTAAC
CTCAATGACTTAGAATGCTTTGCTATATTTTGCCTCTATGCCTCAACCACACTGGCTTTCTTTTAGCTCTTGAACAAGCCA
AACTGCTTCCTGCCTCAGGACCAGATATTTTGGGACTTCTCTTAAGAATTCTATTTCCTTAATTCTTTATCTGGGTAACTTA
GTTTTATCCAACACTTCAGATCCTGCCGTAAAAACTCTTCTTATAGAAGCCTGTCATGACACTGTCTCTCTTCTCCAACATA
CTCACCAGCACACATGTAGACTAGATTAGAACCTCCTGTTTTTCTTTTTCATACTTTTCTCTATCATGCTTCCCTCCATTAT
AATATTTTTATTATGTGTGTGAATGTCTGCCCCAAGTCAGTTTCCTCACTAAACTATAAACTCCGTAAAGCTGGGATCCTTC
CAATTTTGATCACCACTTAGTACAGTAGGAACACAGTAAAGATTCAATTGGTATTTGTGGAATGAATGAATGAATTGTTTTG
CTAGTAAAGTCTGGGGGAACCCAGGTGAGAAGAGCCTAGAAAGCAGGTCGAATCCAAGGCTAGATAGACTTAGTGTTACTCA
AGAAAGGGTAGCCTGAAAATAAAGGTTCAAATTATAGTCAAGAATAGTCAAGACATGGGCAAGACAAGAGTGCTGCTCGTGC
CGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTA
CAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT

SEQ. ID. No. 5
41.1   1288 bp
GAGGGCAGCGAGAAGGAGAAACCCCAGCCCCTGGAGCCCACATCTGCTCTGAGCAATGGGTGCGCCCTCGCCAACCACGCCC
CGGCCCTGCCATGCATCAACCCACTCAGCGCCCTGCAGTCCGTCCTGAACAATCACTTGGGCAAAGCCACGGAGCCCTTGCG
CTCACCTTCCTGCTCCAGCCCAAGTTCAAGCACCAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAGCCGGTC
TTGAGTCCTGCCTCCACAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGAGAACAGCGATCAGCCCATTGACCTGACCA
AGTCCAAAAGCAAGAAAGCCGAGTCCTCGCAAGCACAATCTTGTATGTCCCCACCTCAGAAGCACGCTCTGTCTGACATCGC
CGACATGGTCAAAGTCCTCCCCAAAGCCACCACCCCAAAGCCAGCCTCCTCCTCCAGGGTCCCCCCCATGAAGCTGGAAATG
GATGTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCATAAAAGAAAAGGCCGGCAGTCCAACTGGAATCCTC
AGCATCTTCTGATTCTACAAGCCCAGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGGCAAATACCTGCTGTCTGATCTGGG
CCCACAAGAGCGTATGCAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGCCAACGTCAAGTAC
CAGCTTAGGAAACGGGCGGGACAAAAATTTCTGAAAAACATGGACAAAGGCCACCCCATCTTTTATTGCAGTGACTGTGCCT
CCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTGTC
AGTGGACCAGCAAAGCAAGGTGGAGCAAGAGATCTCCCGGGTATCGTCGGCTCAGAGGTCTCCAGAAACAATAGCTGCCGAA
GAGGACACAGACTCTAAATTCAAGTGTAAGTTGTGCTGTCGGACATTTGTGAGCAAACATGCGGTAAAACTCCACCTAAGCA
AAACGCACAGCAAGTCACCCGAACACCATTCACAGTTTGTAACAGACGTGGATGAAGAATAGCTCTGCAGGACGAATGCCTT
AGTTTCCACTTTCCAGCCTGGATCCCCTCACACTGAACCCTTCTTCGTTGCACCATCCTGCTTCTGACATTGAACTCATTGA
ACTCCTCCTGACACCCTGGCTCTGAGAAGACTGCCAAAAAAAAAAAAAAAAAAAAATTC

SEQ. ID. No. 6
GCAP 2820 bp

```
ATCCTAAGACGCACAGCCTGGGAAGCCAGCACTGGGGAAGTGGTGCTGAGGGATGTGGGTCACTGGGGTGAAGGTGGAGCTT
TCAGGGTCTCCCGTCAATGCAGCTGAGTTTTCTTTGGCAGGGAATTTACCAGCTGAAGAAAGCCTGCCGGCGAGAGCTACAA
ACTGAGCAAGGCCAGCTGCTCACACCCGAGGAGGTCGTGGACAGGATCTTCCTCCTGGTGGATGAGAATGGAGATGGTAAGA
GGGGCAGAGATGGGGAGAGTGCTGTCCACTCTGCATCATCGCCACTTTCTGGCCGCACGTCCTTGGGCAAGGCCCTCCACCT
TCCAACCCTGGGGTCCTCATCTGTGAGAAGGCTGTGGAGAAGATGTCATGAACTAACAAAGGGACTCATGAGCACGTGTTTG
TAGGAGTGACTAAAAGTCCTACAGGAGTTGCTGATGGAGGCCAGGCACGCAGAATAGAAAGAATAGGAACTTTGGAGTCAGG
CAGGGAGTGATATATTGAGCTTCTCGTCCTAGTCTCAATTTCCTCATCTGGAAAATGGGGATAATAATAGTGGTTGAGAGGA
ATGAATAGGATAATGTGTTTAAGAGCAGGCATAGGGTAGACCTCCATTCAGGCTGCTTGGGCTTTCCTCCCTGTAGCCCAAA
GCCCAGCCTCAGGGCTATGTGGGGAGAGAGCTGGCTTGGAATACACACTTGAGCCCTCCAGCTCTCTCAGCTCCACCCAGCA
TTTCCGTGGTACCATGCGCAAAAGTAAAACTTCAATTCATCAGCAAAGAAAGCCCCTTAAAGGTGGCAGGAGACTCCTGGAG
ATTCAGACACCTGACAAGCCGCAAGCTTGAGGTCTGAGACTGCAGGATAGTTGGCATAAGACGTGTAGGCGCATCCTGGGAG
CGAGGTCTCTCCTCCTGCCCCCAGACCCAGGTCTCCCCTTCTTCTACATGACCACCTCTCCTCCCCCTTGCTCAGGCCAGCT
GTCTCTGAACGAGTTTGTTGAAGGTGCCCGTCGGGACAAGTGGGTGATGAAGATGCTGCAGATGGACATGAATCCCAGCAGC
TGGCTCGCTCAGCAGAGACGGAAAAGTGCCATGTTCTGAGGAGTCTGGGGCCCCTCCACGACTCCAGGCTCACCCAGGTTTC
CAGGGTAGTAGGAGGGTCCCCTGGCTCAGCCTGCTCATGCCCACTCTTCCCCTGGTGTTGACTTCCTGGCACCCCCTGTGCA
GGGCTGAGTGGGGATGGGGAAGGGCTGCTGGGTTTGAAGTGGCCAACAGGGCATAGTCCATTTTGGAGGAGTCCCTGGGATG
GTGAAGGGAATTCAGTTACTTTTCCTGTTCAGCCGCTCCTGGGAGGACTGTGCCTTGGCTGGGTGGTTGTGGGGCTCCCACA
GTTTCTGGGTGTTCTCAGTTGGAAGCAAGAGCCAACTGAGGGGTGAGGGTCCCACAGACCAAATCAGAAATGAGAACACAAA
GACTGGTAGGAGGCAGGGGTGGGAGGGTGTTGAGACTGAAGAAAAGGCAGGAGTTGCCGGGCACGGTGGCTCACGCCTGTAA
TCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGGGTGAAACCC
CGTCTCTACTAAAAATACAAAAAATCAGCCGGGTGAGGTGGCGGGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAAG
AGAATGGCGTGAACCCCAGGGGGCCGAGCCTACAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGACGACAGTGAGAC
TCCGTCTCAAAAAAAAAAAAGAAAGAAAAGAAAAAGGCAGGAGTTTTGGGGGGCAGGGGGCAGCAATAATTCTATAACTTCC.
GGGATGCTGAGGGGCGTTCATGGGGAGGACCCTGGCCTCCTCCTCCCCAAGGCATCCTCACCAGTGGTGTCAACAGGAAAAA
TGGCAGCAAATACGCTGCAGGCTGTGGTCTTTCTGCCTTTGAAAGGGTCAGCTGTACTTAAAGGGACTGTTTCAGCTCTGCC
TGGGTGCTGCTCTGGGACCCCCTGCTGCCAACCCACCACTCCCCCAACAATCCTCTCTTTCCATCCATATCCCCCAGTATGG
ACCTTCCACAACTCCCAGCCATAAGCTGAATGTTTCTCTTTAAAGGATGGAGAAAACTTCTGTCTGTCTCTGGCAAGAATTG
GGGGACTGTTGACTGGGATTGTGGGCTGGGCTTGGCTTCTAACTGCTGTGTGACCCAAGACAGCCACTTCTCCTCCCTAACC
TTGGTTATGTCTTGGCAGCACAGTGAGCAGGTCGGACTAGGCGAACAGTTTTGGATTATTGTGTTTTTAGATGTGGAATTAT
TTTTTGTTATATAAACTCTTATGTGTAACCCCAATATAGAAACTAGATTAAAAGGGAGTCTCTCTGGTTGAAAGGGGAGCTG
AGTACCCTCTGGAACTGGAGGCACCTCTGAAAAAAGCAAACTGAAAACCAGTGCCCTGGGTCACTGTTACTCCTATAAGACA
GTTTAAAGTGAGACCTGGAAAAACATTTGCTTTACCTTGAATAGATAGGTTTTTATGTTGGTATATAAGAAATAAAACTAAC
CTATTAACCCTGAGACTTTACAGGTGTGTTATTTCATATGATAGTCATATAAAATTTCCTTTAGACATCAATTTTAGGTAAA
AAATAATTGATTAGAAAAATATTGGCCAGGTGCAGCAGCTCACACCTGCAATCCCAGGACTTTGGGAGGCCGAGGCGGGTGG
ATCACCTGAGGTCAGGGGTTCAAGACCAGCCTG
```

SEQ. ID. No. 7
1b4   1205 bp

```
GCGCGCGTGAGTCCGCCCCCCCAGTCACGTGACCGCTGACTCGGGGCGTTCTCCACTATCGCTTACCTACCTCCCTCTGCAG
GAACCCGGCGATATGGCTGCCGCTGTGCCCCGCGCCGCATTTCTCTCCCCGCTGCTTCCCTTCTCCTGGGCTTCCTGCTCCT
CTCCGCTCCGCATGGCGGCAGCGGCCTGCACACCAAGGCGCCCTTCCCCTGGATACGGTCACTTTCTACAAGGTCATTCCCA
AAAGCAAGTTCGTCTGGTGAAGTTCGACACCCAGTACCCCTACGGTGAGAAGCAGGATGAGTTCAAGCGTCTTCTGAAAAACT
CGGCTTCCAGCGATGATCTCTTGGTGGCAGAGGTGGGGATCTCAGATTATGTGACAAGCTGAACATGGAGCTGAGTGAGAAA
TACAAGCTGGACAAAGAGAGCTACCCATCTTCTACCTCTTCCGGGATGGGGACTTTGAGGAACCCAGTCCCATACACTGGGGC
AGTTAGGTTGGAGCCATCCAGCGCTGGCTGAAGGGCCAAGGGGTCTACCTAGGTATGCCTGGTGCCTGCCTGTATACGACGC
CCTGGCCGGGGAGTTCATCAGGGCCTCTGGTGTGGAGGCCGCCAGGCCCTCTTGAAGCAGGGGCAAGATAACCTCTCAAGTG
TGAAGGAGACTCAGAAGAGTGGGCCGAGCAATACCTGAAGATCATGGGGAAGATCTTAGACCAAGGGGAGCACTTCCAGCAT
CAGAGATGACACGGATCGCCAGGCTGATTGAGAAGAACAAGATGAGTGACGGCAGAAGGAGGAGCTCCAGAAGAGCTTAAAC
ATCCTGACTGCCTTCCAGAAGAAGGGGGCCGAGAAAGAGGAGCTGTAAAAAGGCTGTCTGTGATTTTCCAGGGTTTGGTGGG
GGTAGGGAGGGGANAGTTAACCTGCTGGCTGTGANTCCCTTGTGGAATATAAGGGGGYMSKGGGAAAAGWGGTACTAACCCA
CGATTCTGAGCCCTGAGTATGCCTGGACATTGATGCTAACATGACCATGCTTGGGATGTCTCTAGCTGGTCTGGGGATAGCT
GGAGCACTTACTCAGGTGGCTGGTGAAATGACACCTCAGAAGGAATGAGTGCTATAGAGAGGAGAGAGGAGTGTACTGCCCA
GGTCTTTGACAGATGTAATTCTCATTCAATTAAAGTTTCAGTGTTTTGGTTAAGTGG
```

SEQ. ID. No. 8
20sa7   456 bp

```
GAAATCAGAAGTTTAATATGACACAATTAAATATATTTGTATATCTCACACCGGAGNTTCTCTTCAAACATAAGGAGTTAGA
AATTACAAGTAGGCATATGCTTCCTATATTCAGATAAATTCATTTCGATTAATTAAATTCCAGATAGAGAGAAGTAATTTTC
GGAAAAGAAATGATAGCTATATTAAAGCAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCATCATTCT
```

GAGGCATTTTGGTTGTCATACTGTGTGTGTGGGTGTGCCTGCTGGCATCCAGTGGGCAGAGGCCAGGGACACTGCT
CAGCATGGTACAGTGCACAGGACAGCCCCATCATCAAAGAATTATCTGGTCCCAAATGTCAATAGTTTGAGCATTG
AGAGACCCTAGCCTTCACTTAAGTTTTTCTGGCGTTCCTGATCTTTTTCTGTAGTGAATTCTAGTGGCCATAAAA
GGTACTGGGAGTGATCAACTAGAGCCAGGAATATTATTTGGGCAGCCGTTTGGTGCTGTCCAAAACCTTGTCCTTT

5   CTGTCTGGCAAGCTAGTATCCATTTATAGGTACCTCAGGAACCCAAATGATTTGTCATAAAATACAAGGAATGTGA
GCACACTGAAGACATTTTTAAGAAGGCTCATTTGCTCAGCAGAATTTTCAGTGTACTAGTGGCATTTATAGAAAGA
GAAGGTGATCACTGAAGGCATGCTCACATAATATTCCTGAGCCCTGGTGGGCGTTATCTAGGGCAAAGGATTCCAC
CTGTGTTTGGAGTTGCGCCCATCCTCACTGTAGCCAGAGCTTCTCCTATCAGAGTTTAGTATTTTGTTTGAATAGA
GGATCTTGCTGCTTAAAACAGTTGAAAAGACCCTGATGGGCAGGCCGTAATTGACAAGCGAATGATGGGAACATGA

10   ATCGGTCTTAGGGAAGCATCTGTCAAAGTGGTCCTTGGTTAAAACAAGTGCCTCCTCCTCTCAGTGTCACTTGATT
GTGTGCTTGAATTCTTCGGAAAACTGGGTGTATGAGACCCACGATGAATTTGCCCACACGATTGATTGGACTCTTC
CTTCACCTGCTCTTCAGCCAGTGCCAGTTCCTTTTCTGATCATGTGATTGACGTGAGAACTGTAGTCTGTATATCA
AATCTTTAGAATGTTTTTGAGTTTCCTGGGACACAGGAAACCCAGCACTTAGCATACTACAAATCTAATGTCTTAA
TGGCATCATAAAAAGAGGCTTTAAACACAGACTCCAGTTAGCTAAGTGGTTTCTGCTAGTGCCGGTACTGTTGCAG

15   GGGCCCTGTGAGATGCCCCAGTTCCCTGAAAGAAATGAAAAGGCCAGTTACCGGTAGGTGGTGTGGAAAACATGGG
CTAGATCATCAGGCAGGACAGAATGCCTGGCTGTGGGTGGGAGCACCCCAGCTTGGCGTTGAGTTCTGGTTCTACC
ACTGCGTTGTTTTGTGACCAATTATGAGTTGCTTAACCTTTCTTTGCTACTATTTCCCTGTTTGCAAAATGGTTCA
TTGACCCCTGTCTTCCACCTCCCAAGGACAATTTCAACAGCCTATTTGTAAAAAGATCACAGTCCTTTAAAAAATA
TAACTGTAAAGTCAGAGGTGATGCTTGAAAGAGCAGGAACCAGGTAGATGTGGAAATGTCATGTCCTTTGTTCTAA

20   AGAAAAGGCATTTCATAGCTTTTTGGATATGACGCAACATACCATAAATCCTGACACATAGTTGGGAGTCGGAAAT
TGCAACAACGCCCAGTTATAAACCCAGCTAGTTTGGGTATGATTGTAAGAAAAAAAGCTGGCCATTCTGTATTTG
GGGAATTGATTTTCCTAAACTTATATTATCTTAGTAGTCTAGATTTATCATATTGTACTATCATCCTGGCTTTTTT
AAGACTTAAGAAGATCAAGTAAATTTTTTTTCTTTCTTTAGACACTATATAGATCATCAAGGGTGTCTGTCTTAC
AGGTGGATAGTGATATGATCTACAGTGAGGGGACATTTATTTAAAAACTTAAACATTCATGTGTTTTGGGGGTGGTA

25   TTTTAACGGCAGCACCTCTGATTGTCTTTTGGAGGGCTGGTGTGTGTTTGAAGTTCTGTCCTCCTTCCAGTGGACT
CTAACTTCTCCTGATGCACGTGAGACACATTGTCCTATTGTCCTGCAGAAACTAAAGCCAAACACTGTCATCTGGG
GACAGGTTTTCATTTGTCAGATCTCTTTCGCCCACATGAGTGTTTGTGGACAATACAGCCTGCTTTCCAAAACTTT
GCTAAATTTTGACAGACTTTCCTAGGTGCTTGCCCAATGCCAGACTTTCTTTTCTGTTGAAGATTAAGTTGTGCTT
GCTGCCCTCTAGTGGTCAGTTGTTTAATCCTAACCTTAAACGGCTTATTTTTCCCTGGTGGTTGGGAAGTTGACG

30   GTTTGTAATTGGCTCATTTTTCTAAATTATTCTGAAGAAGATAATTTTTCCCGCCAGTATGTATGTCCACCTTCAG
TTTGCCAGATCCTGCCTGCTCAGAGACACTGAGAACCGGAAGCTGCCCGGGCAATTCAGTCTATGAAATGATCTTT
CTTGTGATTAAGGCAAACGAAGAACTGAATGTTTAATAGTGTACTCTGCTGTACCCAGAAAAAAACAAAACAAAAT
CATGTTATAACACTCTAAAACTTCAAACAACCTCCAACAGCATTTGGTGTGTGTCTAGCCGTTTTGTTCTAACCCG
ATGTTATATAAAGAATTTTTTCATGCTTTCCAAAAATGTTTATGTCAAGAATATTTAAGTCAGCATGCCTTATTC

35   AGGTACTTCAGCTACCTTCTTATATAAATATTTTTGTTTTTCCTTTAAGATAAAAATGATGATGGAGGAAAAATAA
AACATCTTACATCTTCAAGAGAGTGTAGTTATTGTGGAAAGTTTTTCCGTTCAAATTATTACCTCAATATTCATCT
CAGAACGCATACAGGTAAAGAACTTTTATTTTTTTAACCATGCATTAGTTAAATTATGTAGTTATCTAATTTTTTT
GTTGTTGTTGTTCAGATACTCTGCCAGATCCTTGGACTAGCTTAAGGATAAATATGTAGCATGTTGATTGCAGTGG
TTATTTTTATTCTTTTAGTGCCATTGTAACTTGAGCCATTGTTCTTATTTGCAGTTCATTTCTTTTCTTTCTTTTT

40   TGTTTTTTGAGACGGAGTCTTGCTCTGTCACCTCGGCTGGAGTGCAGTGGTGCAATTTCGGCTCACTGCAGCCTCC
ACCTCCCTGGTTCAAGCAATACTCCTGCCTCAGCCTCCCCAGTAGTTGGGATTACAGGTACCTGCCACCACACCCG
GCTAATTTCTGTATTTTTAGTAGAGATGGGGTTTCACCATGCTGGCCAGGCTGGTTTCGAACTCCTGACCTCAAGT
GATCCGCTCACCTTGGCCTCCCATAGTGTTGGCCTCCCATAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGG
ACAAAGTTCATTTGTTTAGTTTATGACTGCTATGTCCTGACTCTTATCTTATTAAAAGCTACAGTATTTTAAAATG

45   CTGCATCTTATGTCTTTATGATTGAGAATGAAATGAGAATCTATTTAGTAGTCTTGAGATTGTGAAAGGAGCTATG
ACATCATGATGTAGGAGGCTGCGTAGATTTGAAATTTCATCTCTTCCACTTACTATCTGTGCACCCTTGGGCAAGT
TATTTAACCTTTTTGTGCTTTTAGTTTTCTTTGCTGTAAAAGTAGAATAATACATATTTCCCTAGGGCTGTTAGGA
AGATTAAATAAGTTAGAAGTGTTGCTGTTAATTTTTCTATTGAAGATAGGCATTCATAATTTCAAATATTCATTAC
AGTAAGGATGATAAAGAACTGATGAGAAATCCTATGTGATAGTAGATCGAGAAAGCAAAAGGAGGAAAGAAGCCTG

50   TTTTCTTAATAAATAGATATTTGATCTATTTCAGTGCTTTTCATACACTTCTATAATAAAGTGCCATTTCTTGCCT
TAGGTGAAAAAACCATACAAATGTGAATTTTGTGAATATGCTGCAGCCCAGAAGACATCTCTGAGGTATCACTTGGA
GAGACATCACAAGGAAAAACAAACCGATGTTGCTGCTGAAGTCAAGAACGATGGTAAAAATCAGGACACTGAAGAT
GCACTATTAACCGCTGACAGTGCGCAAACCAAAAATTTGAAAAGATTTTTTGATGGTGCCAAAGATGTTACAGGCA
GTCCACCTGCAAAGCAGCTTAAGGAGATGCCTTCTGTTTTTCAGAATGTTCTGGGCAGCGCTGTCCTCTCACCAGC

55   ACACAAAGATACTCAGGATTTCCATAAAAATGCAGCTGATGACAGTGCTGATAAAGTGAATAAAAACCCTACCCCT
GCTTACCTGGACCTGTTAAAAAAGAGATCAGCAGTTGAAACTCAGGCAAATAACCTCATCTGTAGAACCAAGGCGG
ATGTTACTCCTCCTCCGGATGGCAGTACCACCCATAACCTTGAAGTTAGCCCCAAAGAGAAGCAAACGGAGACCGC
AGCTGACTGCAGATACAGGCCAAGTGTGGATTGTCACGAAAAACCTTTAAATTTATCCGTGGGGGCTCTTCACAAT
TGCCCGGCAATTTCTTTGAGTAAAAGTTTGATTCCAAGTATCACCTGTCCATTTTGTACCTTCAAGACATTTTATC

60   CAGAAGTTTTAATGATGCACCAGAGACTCGAGCATAAATACAATCCTGACGTTCATAAAAACTGTCGAAACAAGTC
CTTGCTTAGAAGTCGACGTACCGGATGCCCGCCAGCGTTGCTGGGAAAGATGTGCCTCCCCTCCCTAGTTTCTGT
AAACCCAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAAATCCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTCCTG
GGCCAGGCAAGGCCCCTCTGACTTCAGGGATAGACTCTAGCACTTTAGCCCCAAGTAACCTGAAGTCCCACAGACC
ACAGCAGAATGTGGGGGTCCAAGGGGCCGCCACCAGGCAACAGCAATCTGAGATGTTTCCTAAAACCAGTGTTTCC

65   CCTGCACCGGATAAGACAAAAAGACCCGAGACAAAATTGAAACCTCTTCCAGTAGCTCCTTCTCAGCCCACCCTCG
GCAGCAGTAACATCAATGGTTCCATCGACTACCCGCCAAGAACGACAGCCCGTGGGCACCTCCGGGAAGAGACTA
TTTCTGTAATCGGAGTGCCAGCAATACTGCAGCAGAATTTGGTGAGCCCCTTCCAAAAAGACTGAAGTCCAGCGTG

GTTGCCCTTGACGTTGACCAGCCCGGGGCCAATTACAGAAGAGGCTATGACCTTCCCAAGTACCATATGGTCAGAG
GCATCACATCACTGTTACCGCAGGACTGTGTGTATCCGTCGCAGGCGCTGCCTCCCAAACCAAGGTTCCTGAGCTC
CAGCGAGGTCGATTCTCCAAATGTGCTGACTGTTCAGAAGCCCTATGGTGGCTCCGGGCCACTTTACACTTGTGTG
CCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAAGGTATTGCATGAGGGGCGTCGTGTTTAAATGGCTGCCTAC
AGTGATTAATAGCTAATCCAGGCATTCTCAGTGGAGATGGTACCACTCCCAAGGGTGGGGGGTAGGCAGCCAGAAG
TTCTTGGGGGTCACAGAGAGAAGCATTCTTAGATACGGCAGTGGTTTGTGGTCCTCCAAGGCTTACTTAACTCTGT
GGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTT
GCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATC
CTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTG
TAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCT
TAGACTACAGATGCACTCCACCACGCTTG

SEQ. ID. No. 10
ZABC1 Open reading frame

ATGCAATCGAAAGTGACAGGAAACATGCCAACTCAATCCCTCTTAATGTACATGGATGGGCCAGAAGTGATTGGCA
GCTCTCTTGGCAGTCCGATGGAGATGGAGGATGCCTTGTCAATGAAAGGGACCGCTGTTGTTCCATTCCGAGCTAC
ACAAGAAAAAAATGTCATCCAAATCGAGGGGTATATGCCCTTGGATTGCATGTTCTGCAGCCAGACCTTCACACAT
TCAGAAGACCTTAATAAACATGTCTTAATGCAACACCGGCCTACCCTCTGTGAACCAGCAGTTCTTCGGGTTGAAG
CAGAGTATCTCAGTCCGCTTGATAAAAGTCAAGTGCGAACAGAACCTCCCAAGGAAAAGAATTGCAAGGAAAATGA
ATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTCGCTTTTGATGTTGAGATCCACATGAGAACACACAAAGAT
TCTTTCACTTACGGGTGTAACATGTGCGGAAGAAGMTTSRRSSAGCCTTGGTTTCTTAAAAATCACATGCGGACAC
ATAATGGCAAATCGGGGGCCAGAAGCAAACTGCAGCAAGGCTTGGAGAGTAGTCCAGCAACGATCAACGAGGTCGT
CCAGGTGCACGCGGCCGAGAGCATCTCCTCTCCTTACAAAATCTGCATGGTTTGTGGCTTCCTATTTCCAAATAAA
GAAAGTCTAATTGAGCACCGCCAAGGTGCACACCAAAAAAAACTGCTTTCGGTACCAGCAGCGCGCAGACAGACTCTC
CACAAGGAGGAATGCCGTCCTCGAGGGAGGACTTCCTGCAGTTGTTCAACTTGAGACCAAAATCTCACCCTGAAAC
GGGGAAGAAGCCTGTCAGATGCATCCCTCAGCTCGATCCGTTCACCACCTTCCAGGCTTGGCAGCTGGCTACCAAA
GGAAAAGTTGCCATTTGCCAAGAAGTGAAGGAATCGGGGCAAGAAGGGAGCACCGACAACGACGATTCGAGTTCCG
AGAAGGAGCTTGGAGAAACAAATAAGGGCAGTTGTGCAGGCCTCTCGCAAGAGAAAGAGAAGTGCAAACACTCCCA
CGGCGAAGCGCCCTCCGTGGACGCGGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTCACTGCTCCGAGTGC
GGCAAAGCTTTCAGAACCTACCACCAGCTGGTCTTGCACTCCAGGGTCCACAAGAAGGACCGGAGGGCCGGCGCGG
AGTCGCCCACCATGTCTGTGGACGGGAGGCAGCCGGGGACGTGTTCTCCTGACCTCGCCGCCCCTCTGGATGAAAA
TGGAGCCGTGGATCGAGGGGAAGGTGGTTCTGAAGACGGATCTGAGGATGGGCTTCCCGAAGGAATCCATCTGGAT
AAAAATGATGATGGAGGAAAAATAAAACATCTTACATCTTCAAGAGAGTGTAGTTATTGTGGAAAGTTTTTCCGTT
CAAATTATTACCTCAATATTCATCTCAGAACGCATACAGGTGAAAAACCATACAAATGTGAATTTTGTGAATATGC
TGCAGCCCAGAAGACATCTCTGAGGTATCACTTGGAGAGACATCACAAGGAAAAACAAACCGATGTTGCTGCTGAA
GTCAAGAACGATGGTAAAAATCAGGACACTGAAGATGCACTATTAACCGCTGACAGTGCGCAAACCAAAAATTTGA
AAAGATTTTTTGATGGTGCCAAAGATGTTACAGGCAGTCCACCTGCAAAGCAGCTTAAGGAGATGCCTTCTGTTTT
TCAGAATGTTCTGGGCAGCGCTGTCCTCTCACCAGCACACAAAGATACTCAGGATTTCCATAAAAATGCAGCTGAT
GACAGTGCTGATAAAGTGAATAAAAACCCTACCCCTGCTTACCTGGACCTGTTAAAAAAGAGATCAGCAGTTGAAA
CTCAGGCAAATAACCTCATCGTAGAACCAAGGCGGATGTTACTCCTCCTCCGGATGGCAGTACCACCCATAACCT
TGAAGTTAGCCCCAAAGAGAAGCAAACGGAGACCGCAGCTGACTGCAGATACAGGCCAAGTGTGGATTGTCACGAA
AAACCTTTAAATTTATCCGTGGGGGCTCTTCACAATTGCCCGGCAATTTCTTTGAGTAAAAGTTTGATTCCAAGTA
TCACCTGTCCATTTTGTACCTTCAAGACATTTATCCAGAAGTTTTAATGATGCACCAGAGACTGGAGCATAAATA
CAATCCTGACGTTCATAAAAACTGTCGAAACAAGTCCTTGCTTAGAAGTCGACGTACCGGATGCCCGCCAGCGTTG
CTGGGAAAGATGTGCCTCCCCTCTCTAGTTTCTGTAAACCCAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAAAT
CCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTCCTGGGCCAGGCAAGGCCCCTCTGACTTCAGGGATAGACTCTAG
CACTTTAGCCCCAAGTAACCTGAAGTCCCACAGACCACAGCAGAATGTGGGGGTCCAAGGGGCCGCCACCAGGCAA
CAGCAATCTGAGATGTTTCCTAAAACCAGTGTTTCCCTGCACCGGATAAGACAAAAAGACCCGAGACAAAATTGA
AACCTCTTCCAGTAGCTCCTTCTCAGCCCACCCTCGGCAGCAGTAACATCAATGGTTCCATCGACTACCCCGCCAA
GAACGACAGCCCGTGGGCACCTCCGGGAAGAGACTATTTCTGTAATCGGAGTGCCAGCAATACTGCAGCAGAATTT
GGTGAGCCCCTTCCAAAAAGACTGAAGTCCAGCGTGGTTGCCCTTGACGTTGACCAGCCCGGGGCCAATTACAGAA
GAGGCTATGACCTTCCCAAGTACCATATGGTCAGAGGCATCACATCACTGTTACCGCAGGACTGTGTGTATCCGTC
GCAGGCGCTGCCTCCCAAACCAAGGTTCCTGAGCTCCAGCGAGGTCGATTCTCCAAATGTGCTGACTGTTCAGAAG
CCCTATGGTGGCTCCGGGCCACTTTACACTTGTGTGCCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAAGGTC
TTGGTGGATGTCAGTGCTTACTCCCCATGAAATTAAATTTTACTTCATCCTTTGAGAAGCGAATGGTGAAAGCTAC
TGAAATAAGCTGTGATTGTACTGTACATAAAACATATGAGGAATCTGCAAGGAACACTACAGTTGTGTAA

SEQ. ID. No. 11
ZABC1 Protein

MQSKVTGNMPTQSLLMYMDGPEVIGSSLGSPMEMEDALSMKGTAVVPFRATQEKNVIQIEGYMPLDCMFCSQTFTH
SEDLNKHVLMQHRPTLCEPAVLRVEAEYLSPLDKSQVRTEPPKEKNCKENEFSCEVCGQTFRVAFDVEIHMRTHKD
SFTYGCNMCGRXXXXPWFLKNHMRTHNGKSGARSKLQQGLESSPATINEVVQVHAAESISSPYKICMVCGFLFPNK
ESLIEHRKVHTKKTAFGTSSAQTDSPQGGMPSSREDFLQLFNLRPKSHPETGKKPVRCIPQLDPFTTFQAWQLATK
GKVAICQEVKESGQEGSTDNDDSSSEKELGETNKGSCAGLSQEKEKCKHSHGEAPSVDADPKLPSSKEKPTHCSEC
GKAFRTYHQLVLHSRVHKKDRRAGAESPTMSVDGRQPGTCSPDLAAPLDENGAVDRGEGGSEDGSEDGLPEGIHLD

KNDDGGKIKHLTSSRECSYCGKFFRSNYYLNIHLRTHTGEKPYKCEFCEYAAAQKTSLRYHLERHHKEKQTDVAAE
VKNDGKNQDTEDALLTADSAQTKNLKRFFDGAKDVTGSPPAKQLKEMPSVFQNVLGSAVLSPAHKDTQDFHKNAAD
DSADKVNKNPTPAYLDLLKKRSAVETQANNLICRTKADVTPPPDGSTTHNLEVSPKEKQTETAADCRYRPSVDCHE
KPLNLSVGALHNCPAISLSKSLIPSITCPFCTFKTFYPEVLMMHQRLEHKYNPDVHKNCRNKSLLRSRRTGCPPAL
LGKDVPPLSSFCKPKPKSAFPAQSKSLPSAKGKQSPPGPGKAPLTSGIDSSTLAPSNLKSHRPQQNVGVQGAATRQ
QQSEMFPKTSVSPAPDKTKRPETKLKPLPVAPSQPTLGSSNINGSIDYPAKNDSPWAPPGRDYFCNRSASNTAAEF
GEPLPKRLKSSVVALDVDQPGANYRRGYDLPKYHMVRGITSLLPQDCVYPSQALPPKPRFLSSSEVDSPNVLTVQK
PYGGSGPLYTCVPAGSPASSSTLEGLGGCQCLLPMKLNFTSSFEKRMVKATEISCDCTVHKTYEESARNTTVV

SEQ. ID. NO. 12

1b1

GGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACC
GCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACGAGGCTCCACCGACAGCCAGG
CACTGGGCAGCACGCACTGGAGACCCAGGACCCTGTGCAGGAGCAGCTCCGGGTGACACGAGGGGACTGAAGATAC
TCCCACAGGGGCTCAGCAGGAGCAATGGGTAACCAAATGAGTGTTCCCCAAAGAGTTGAAGACCAAGAGAATGAAC
CAGAAGCAGAGACTTACCAGGACAACGCGTCTGCTCTGAACGGGGTTCCAGTGGTGGTGTCGACCCACACAGTTCA
GCACTTAGAGGAAGTCGACTTGGGAATAAGTGTCAAGACGGATAATGTGGCCACTTCTTCCCCCGAGACAACGGAG
ATAAGTGCTGTTGCGCATGCCAACGGAAAGAATCTTGGGAAAGAGGCCAAACCCGAGGCACCAGCTGCTAAATCTC
GTTTTTTCTTGATGCTCTCTCGGCCTGTACCAGGACGTACCGGAGACCAAGCCGCAGATTCATCCCTTGGATCAGT
GAAGCTTGATGTCAGCTCCAATAAAGCTCCAGCGAACAAAGACCCAAGTGAGAGCTGGACACTTCCGGTGGCAGCT
GGACCGGGGCAGGACACAGATAAAACCCCAGGGCACGCCCCGGCCCAAGACAAGGTCCTCTCTGCCGCCAGGGATC
CCACGCTTCTCCCACCTGAGACAGGGGGAGCAGGAGGAGAAGCTCCCTCCAAGCCCAAGGACTCCAGCTTTTTTGA
CAAATTCTTCAAGCTGGACAAAGGGACAGGAAAAGGTGCCAGGTGACAGCCAACAGGAAGCCAAGAGGGCAGAGCAT
CAAGACAAGGTGGATGAGGTTCCTGGCTTATCAGGGCAGTCCGATGATGTCCCTGCAGGGAAGGACATAGTTGACG
GCAAGGAAAAGAAGGACAAGAACTTGGAACTGCGGATTGCTCTGTCCCTGGGGACCCAGAAGGACTGGAGACTGC
AAAGGACGATTCCCAGGCAGCAGCTATAGCAGAGAATAATAATTCCATCATGAGTTTCTTTAAAACTCTGGTTTCA
CCTAACAAAGCTGAAACAAAAAAGGACCCAGAAGACACGGGTGCTGAAAAGTCACCCACCACTTCAGCTGACCTTA
AGTCAGACAAAGCCAACTTTACATCCCAGGAGACCCAAGGGGCTGGCAAGAATTCCAAAGGATGCAACCCATCGGG
GCACACAGTCCGTGACAACCCCTGAACCTGCGAAGGAAGGCACCAAGGAGAAATCAGGACCCACCTCTCTGCCT
CTGGGCAAACTGTTTTGGAAAAAGTCAGTTAAAGAGGACTCAGTCCCCACAGGTGCGGAGGAGAATGTGGTGTGTG
AGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGGAGATGC
TGCACCTGAACCCACAGAAGCGAAACTCAAAAGAGAAGAAAGCAAACCAAGAACCTCTCTGATGGCGTTTCTCAGA
CAAATGTCAGTGAAAGGGGATGGAGGGATCACCCACTCAGAAGAAATAAATGGGAAAGACTCCAGCTGCCAAACAT
CAGACTCCACAGAAAAGACTATCACACCGCCAGAGCCTGAACCAACAGGAGCACCACAGAAGGGTAAAGAGGGCTC
CTCGAAGGACAAGAAGTCAGCAGCCGAGATGAACAAGCAGAAGAGCAACAAGCAGGAAGCCAAAGAACCAGCCCAG
TGCACAGAGCAGGCCACGGTGGACACGAACTCACTGCAGAATGGGGACAAGCTCCAAAAGAGACCTGAGAAGCGGC
AGCAGTCCCTTGGGGGCTTCTTTAAAGGCCTGGGACCAAAGCGGATGTTGGATGCTCAAGTGCAAACAGACCCAGT
ATCCATCGGACCAGTTGGCAAACCCAAGTAAACAAATCAGCACGGTTCCCACCAGGTTCTCCTGCCACCAAGATGT
GTTCTCCTTACTCCATCTCCTCCCCAAACACGCTCCATGTATATATTCTTCTGATGGCCAGCAAATGAAATTCTGC
CTAGAAATTAAGCCCGAGCTGTTGTATATTGAGGTGTATTATTTACGTCTCTGGTCCAGTCTTTTCTGGCAAATAA
CAGTAAAGATGGTTTAGCAGGTCACCTAGTTGGGTCAGAAGAGTCGATGATCACCAAGCAGGAAAGGGAGGGAATA
GAGGAATGTGTTCGGGTTAAGTGATGAAAATGGCAGTGGTGGCCGGGCGTGGTGGCTCTCGCCTGTAATCTCAGCA
CTTTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACTAGCCTGGCCAACATCATGAAACCCCG
TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTGTAGTCCCAGCTACTCGGGAGCCCAACGC
ACGAGAACCGCTTGTACCCAGGAGGTGGAGGTTGCAGTGAGCCGAAGTTGCACCATTGCACTCCACCCTGGGCGAC
AGAGCAAGATTCTATCAAAAAAAAAAGGCAGTGGCAAGTAAGTTATAGAAGAGAAATGCTGCTAGAAGGAATTAAG
CGTTGTAGTAAACGCGTGCTCATCCTCTAAGCTTGAAGAAGGGAGACGAAAATCCATTTGTTTAAATTCACATCTC
AAGGAGGGAGAACCCGGGCTGTGTTGGGTGGTTGCCAATTTCCTAGAACGGAATGTGTGGGGTATAGAAAAAGGAA
TGAATAAGCGTTGTTTTTCAAATAGGGTCCTTGTAAGTTATTGATGAGAGGGAAAGATTGACTGGGGAGGGCTTA
AAATGATTTGGGAAAACAATTGCTTTTGAGGCTCAGTGACAACGGCAAAGATTACAACTTAAAAAAAAAAAAAAAAA
AAACTCGAGACTAGTTCTCTCTCTCTCGTGCCGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGG
GGGCCCGGTACCCAATTCGCCCTATA